A First Course in Linear Algebra

A First Course in Linear Algebra

by

Robert A. Beezer Department of Mathematics and Computer Science University of Puget Sound

Version 2.99

Robert A. Beezer is a Professor of Mathematics at the University of Puget Sound, where he has been on the faculty since 1984. He received a B.S. in Mathematics (with an Emphasis in Computer Science) from the University of Santa Clara in 1978, a M.S. in Statistics from the University of Illinois at Urbana-Champaign in 1982 and a Ph.D. in Mathematics from the University of Illinois at Urbana-Champaign in 1984. He teaches calculus, linear algebra and abstract algebra regularly, while his research interests include the applications of linear algebra to graph theory. His professional website is at http://buzzard.ups.edu.

Edition

Version 2.99. October 13, 2012.

Publisher

Robert A. Beezer Department of Mathematics and Computer Science University of Puget Sound 1500 North Warner Tacoma, Washington 98416-1043 USA

© 2004 by Robert A. Beezer.

Permission is granted to copy, distribute and/or modify this document under the terms of

V

the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the appendix entitled "GNU Free Documentation License".

The most recent version of this work can always be found at http://linear.ups.edu.

To my wife, Pat.

Contents

Table of Contents	vii
Contributors	viii
Definitions	ix
Theorems	x
Notation	xi
Diagrams	xii
Examples	xiii
Preface	xiv

Acknowledgements

Part C Core

Chapter SLE Systems of Linear Equations
WILA What is Linear Algebra?
LA "Linear" + "Algebra"
AA An Application
READ Reading Questions
EXC Exercises
SOL Solutions
SSLE Solving Systems of Linear Equations
SLE Systems of Linear Equations
PSS Possibilities for Solution Sets
ESEO Equivalent Systems and Equation Operations
READ Reading Questions
EXC Exercises
SOL Solutions
RREF Reduced Row-Echelon Form
MVNSE Matrix and Vector Notation for Systems of Equations
RO Row Operations

xxvii

RREF Reduced Row-Echelon Form	63
READ Reading Questions	86
EXC Exercises	88
SOL Solutions	97
TSS Types of Solution Sets	112
CS Consistent Systems	112
FV Free Variables	124
READ Reading Questions	130
EXC Exercises	131
SOL Solutions	138
HSE Homogeneous Systems of Equations	146
SHS Solutions of Homogeneous Systems	146
NSM Null Space of a Matrix	152
READ Reading Questions	157
EXC Exercises	158
SOL Solutions	164
NM Nonsingular Matrices	172
NM Nonsingular Matrices	172
NSNM Null Space of a Nonsingular Matrix	178
READ Reading Questions	182
EXC Exercises	184
SOL Solutions	188

Chapter V Vectors 1	97
VO Vector Operations	98
VEASM Vector Equality, Addition, Scalar Multiplication	.99
VSP Vector Space Properties	205
READ Reading Questions	208
	209
SOL Solutions	215
LC Linear Combinations	221
LC Linear Combinations	221
VFSS Vector Form of Solution Sets	232
PSHS Particular Solutions, Homogeneous Solutions	256
READ Reading Questions	261
EXC Exercises	263
SOL Solutions	267
SS Spanning Sets	270
SSV Span of a Set of Vectors	270
SSNS Spanning Sets of Null Spaces	283
READ Reading Questions 2	293
EXC Exercises	295
SOL Solutions	B 01
LI Linear Independence	816
LISV Linearly Independent Sets of Vectors	B 16
LINM Linear Independence and Nonsingular Matrices	3 28

NSSLI Null Spaces, Spans, Linear Independence	331
READ Reading Questions	337
EXC Exercises	339
SOL Solutions	348
LDS Linear Dependence and Spans	364
LDSS Linearly Dependent Sets and Spans	364
COV Casting Out Vectors	37(
READ Reading Questions	38!
EXC Exercises	386
SOL Solutions	389
O Orthogonality	395
CAV Complex Arithmetic and Vectors	395
IP Inner products	398
N Norm	404
OV Orthogonal Vectors	408
GSP Gram-Schmidt Procedure	413
READ Reading Questions	421
EXC Exercises	422
SOL Solutions	424
Chapter M Matrices	42
MO Matrix Operations	420
MEASM Matrix Equality, Addition, Scalar Multiplication	426
MEASM Matrix Equality, Addition, Scalar Multiplication	42

VSP Vector Space Properties	430
TSM Transposes and Symmetric Matrices	433
MCC Matrices and Complex Conjugation	438
AM Adjoint of a Matrix	442
READ Reading Questions	444
EXC Exercises	446
SOL Solutions	453
MM Matrix Multiplication	460
MVP Matrix-Vector Product	460
MM Matrix Multiplication	468
MMEE Matrix Multiplication, Entry-by-Entry	471
PMM Properties of Matrix Multiplication	474
HM Hermitian Matrices	485
READ Reading Questions	488
EXC Exercises	490
SOL Solutions	496
MISLE Matrix Inverses and Systems of Linear Equations	503
IM Inverse of a Matrix	506
CIM Computing the Inverse of a Matrix	509
PMI Properties of Matrix Inverses	519
READ Reading Questions	524
EXC Exercises	525
SOL Solutions	531

MINM Matrix Inverses and Nonsingular Matrices	537
NMI Nonsingular Matrices are Invertible	537
UM Unitary Matrices	544
READ Reading Questions	552
EXC Exercises	553
SOL Solutions	556
CRS Column and Row Spaces	560
CSSE Column Spaces and Systems of Equations	561
CSSOC Column Space Spanned by Original Columns	567
CSNM Column Space of a Nonsingular Matrix	572
RSM Row Space of a Matrix	576
READ Reading Questions	589
EXC Exercises	591
SOL Solutions	601
FS Four Subsets	609
LNS Left Null Space	609
CCS Computing Column Spaces	612
EEF Extended echelon form	618
FS Four Subsets	624
READ Reading Questions	644
EXC Exercises	645
SOL Solutions	650

Chapter VS Vector Spaces	658
VS Vector Spaces	659
VS Vector Spaces	659
EVS Examples of Vector Spaces	663
VSP Vector Space Properties	674
RD Recycling Definitions	681
READ Reading Questions	682
EXC Exercises	683
SOL Solutions	688
S Subspaces	694
TS Testing Subspaces	698
TSS The Span of a Set	707
SC Subspace Constructions	719
READ Reading Questions	720
EXC Exercises	722
SOL Solutions	726
LISS Linear Independence and Spanning Sets	734
LI Linear Independence	735
SS Spanning Sets	745
VR Vector Representation	754
READ Reading Questions	758
EXC Exercises	759
SOL Solutions	764

B Bases	777
B Bases	777
BSCV Bases for Spans of Column Vectors	785
BNM Bases and Nonsingular Matrices	789
OBC Orthonormal Bases and Coordinates	792
READ Reading Questions	801
EXC Exercises	802
SOL Solutions	805
D Dimension	815
D Dimension	815
DVS Dimension of Vector Spaces	824
RNM Rank and Nullity of a Matrix	828
RNNM Rank and Nullity of a Nonsingular Matrix	832
READ Reading Questions	835
EXC Exercises	837
SOL Solutions	842
PD Properties of Dimension	850
GT Goldilocks' Theorem	850
RT Ranks and Transposes	859
DFS Dimension of Four Subspaces	861
	863
READ Reading Questions	874
	876

SOL Solutions	879
Chapter D Determinants	882
DM Determinant of a Matrix	883
EM Elementary Matrices	883
DD Definition of the Determinant	893
CD Computing Determinants	896
READ Reading Questions	905
EXC Exercises	907
SOL Solutions	912
PDM Properties of Determinants of Matrices	919
DRO Determinants and Row Operations	919
DROEM Determinants, Row Operations, Elementary Matrices	930
DNMMM Determinants, Nonsingular Matrices, Matrix Multiplication .	934
READ Reading Questions	940
EXC Exercises	941
SOL Solutions	943
Chapter E Eigenvalues	944
EE Eigenvalues and Eigenvectors	945
EEM Eigenvalues and Eigenvectors of a Matrix	945
PM Polynomials and Matrices	949
EEE Existence of Eigenvalues and Eigenvectors	952

READ Reading Questions983EXC Exercises984SOL Solutions982PEE Properties of Eigenvalues and Eigenvectors1001ME Multiplicities of Eigenvalues1013EHM Eigenvalues of Hermitian Matrices1002READ Reading Questions1002EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1032D Diagonalization1032D Diagonalization1032EXC Exercises1054READ Reading Questions1059EXC Exercises1054READ Reading Questions1059Chapter LT Linear Transformations1071LT Linear Transformations1072	CEE Computing Eigenvalues and Eigenvectors	960
EXC Exercises984SOL Solutions989PEE Properties of Eigenvalues and Eigenvectors1001ME Multiplicities of Eigenvalues1013EHM Eigenvalues of Hermitian Matrices1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1022PSM Properties of Similar Matrices1032D Diagonalization1032EXC Exercises1044SOL Solutions1026SD Similarity and Diagonalization1029PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072	ECEE Examples of Computing Eigenvalues and Eigenvectors	968
SOL Solutions989PEE Properties of Eigenvalues and Eigenvectors1001ME Multiplicities of Eigenvalues1013EHM Eigenvalues of Hermitian Matrices1020READ Reading Questions1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1032D Diagonalization1032Chapter LT Linear Transformations1071LT Linear Transformations1072	READ Reading Questions	983
PEE Properties of Eigenvalues and Eigenvectors1001ME Multiplicities of Eigenvalues1013EHM Eigenvalues of Hermitian Matrices1020READ Reading Questions1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1032D Diagonalization1032Chapter LT Linear Transformations1061LT Linear Transformations1071	EXC Exercises	984
ME Multiplicities of Eigenvalues1013EHM Eigenvalues of Hermitian Matrices1020READ Reading Questions1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1032D Diagonalization1037FS Fibonacci Sequences1059EXC Exercises1060SOL Solutions1060SOL Solutions1071LT Linear Transformations1072	SOL Solutions	989
EHM Eigenvalues of Hermitian Matrices1020READ Reading Questions1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029SM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1059EXC Exercises1060SOL Solutions1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072	PEE Properties of Eigenvalues and Eigenvectors	1001
READ Reading Questions1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072		1013
READ Reading Questions1023EXC Exercises1024SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072	EHM Eigenvalues of Hermitian Matrices	1020
SOL Solutions1026SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1054READ Reading Questions1059EXC Exercises1060SOL Solutions1063 Chapter LT Linear Transformations LT Linear Transformations1071		1023
SD Similarity and Diagonalization1029SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1054READ Reading Questions1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072	EXC Exercises	1024
SM Similar Matrices1029PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1054READ Reading Questions1059EXC Exercises1060SOL Solutions1063 Chapter LT Linear Transformations LT Linear Transformations1072	SOL Solutions	1026
PSM Properties of Similar Matrices1032D Diagonalization1037FS Fibonacci Sequences1054READ Reading Questions1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072	SD Similarity and Diagonalization	1029
D Diagonalization1037FS Fibonacci Sequences1054READ Reading Questions1059EXC Exercises1060SOL Solutions1063Chapter LT Linear Transformations1071LT Linear Transformations1072		1029
FS Fibonacci Sequences 1054 READ Reading Questions 1059 EXC Exercises 1060 SOL Solutions 1063 Chapter LT Linear Transformations 1071 LT Linear Transformations 1072	PSM Properties of Similar Matrices	1032
READ Reading Questions 1059 EXC Exercises 1060 SOL Solutions 1063 Chapter LT Linear Transformations 1071 LT Linear Transformations 1072	D Diagonalization	1037
EXC Exercises 1060 SOL Solutions 1063 Chapter LT Linear Transformations 1071 LT Linear Transformations 1072	FS Fibonacci Sequences	1054
SOL Solutions 1063 Chapter LT Linear Transformations 1071 LT Linear Transformations 1072	READ Reading Questions	1059
Chapter LT Linear Transformations 1071 LT Linear Transformations 1072	EXC Exercises	1060
LT Linear Transformations	SOL Solutions	1063
	Chapter LT Linear Transformations	1071
	LT Linear Transformations	1072
	LT Linear Transformations	1072

MLT Matrices and Linear Transformations108LTLC Linear Transformations and Linear Combinations109PI Pre-Images110NLTFO New Linear Transformations From Old110)2)1)5
PI Pre-Images)1)5
)5
NLIFO New Linear Hansiormations From Oid	4
READ Reading Questions	· +
EXC Exercises	.5
SOL Solutions $\ldots \ldots \ldots$	20
ILT Injective Linear Transformations	28
EILT Examples of Injective Linear Transformations	29
KLT Kernel of a Linear Transformation	37
ILTLI Injective Linear Transformations and Linear Independence 114	17
ILTD Injective Linear Transformations and Dimension	i 0
CILT Composition of Injective Linear Transformations	52
READ Reading Questions	53
EXC Exercises	j 4
SOL Solutions $\ldots \ldots \ldots$	i 0
SLT Surjective Linear Transformations	;9
ESLT Examples of Surjective Linear Transformations	'0
RLT Range of a Linear Transformation	'8
SSSLT Spanning Sets and Surjective Linear Transformations	38
SLTD Surjective Linear Transformations and Dimension 119)3
CSLT Composition of Surjective Linear Transformations)4

READ Reading Questions	.96
EXC Exercises	97
SOL Solutions	203
IVLT Invertible Linear Transformations 12	213
	213
IV Invertibility	221
	229
RNLT Rank and Nullity of a Linear Transformation	233
SLELT Systems of Linear Equations and Linear Transformations 12	240
	244
EXC Exercises	245
SOL Solutions	251
Chapter R Representations 12	60
	261
	272
	275
	281
	282
	284
	286
*	800
	810

IVLT Invertible Linear Transformations	1321
READ Reading Questions	1330
EXC Exercises	1331
SOL Solutions	1337
CB Change of Basis	1355
	1355
CBM Change-of-Basis Matrix	1358
MRS Matrix Representations and Similarity	1371
	1385
READ Reading Questions	1402
EXC Exercises	1404
	1406
	1414
	1415
	1418
	1427
	1428
	1435
	1436
	1448
	1458
IS Invariant Subspaces	1475
	1475

GEE Generalized Eigenvectors and Eigenspaces	1484
RLT Restrictions of Linear Transformations	1494
EXC Exercises	1516
JCF Jordan Canonical Form	1517
GESD Generalized Eigenspace Decomposition	1518
JCF Jordan Canonical Form	1532
CHT Cayley-Hamilton Theorem	1563
Appendix CN Computation Notes	1565
MMA Mathematica	1566
ME.MMA Matrix Entry	1566
RR.MMA Row Reduce	1567
LS.MMA Linear Solve	1567
VLC.MMA Vector Linear Combinations	1568
NS.MMA Null Space	1569
VFSS.MMA Vector Form of Solution Set	1571
GSP.MMA Gram-Schmidt Procedure	1573
TM.MMA Transpose of a Matrix	1575
MM.MMA Matrix Multiplication	1575
MI.MMA Matrix Inverse	1576
TI86 Texas Instruments 86	1577
ME.TI86 Matrix Entry	1577
RR.TI86 Row Reduce	1578

VLC.TI86 Vector Linear Combinations	79
TM.TI86 Transpose of a Matrix	80
	80
ME.TI83 Matrix Entry	80
RR.TI83 Row Reduce 15	81
VLC.TI83 Vector Linear Combinations	82
SAGE SAGE: Open Source Mathematics Software	83
R.SAGE Rings	83
ME.SAGE Matrix Entry	84
RR.SAGE Row Reduce	86
LS.SAGE Linear Solve	88
VLC.SAGE Vector Linear Combinations	89
MI.SAGE Matrix Inverse	90
TM.SAGE Transpose of a Matrix	91
E.SAGE Eigenspaces	91
Appendix P Preliminaries 15	93
CNO Complex Number Operations	94
CNA Arithmetic with complex numbers	94
	98
	00
SET Sets	02
SC Set Cardinality	06

SO Set Operations	1607
PT Proof Techniques	1611
D Definitions	1611
T Theorems	1613
L Language	1614
GS Getting Started	1617
C Constructive Proofs	1618
E Equivalences	1619
N Negation	1620
CP Contrapositives	1621
CV Converses	1622
CD Contradiction	1623
U Uniqueness	1625
ME Multiple Equivalences	1626
PI Proving Identities	1627
DC Decompositions	1628
I Induction	1630
P Practice	1633
LC Lemmas and Corollaries	1634
ppendix A Archetypes	1637
A	1642
B	1652

\mathbf{C}	 1662
D	 1670
\mathbf{E}	 1678
\mathbf{F}	 1686
G	 1697
Η	 1705
Ι	 1714
J	 1723
Κ	 1734
\mathbf{L}	 1742
Μ	 1750
Ν	 1755
Ο	 1760
Р	 1766
\mathbf{Q}	 1771
R	 1779
\mathbf{S}	 1785
Т	 1790
U	 1794
V	 1799
W	 1803
Х	 1808

Appendix GFDL GNU Free Documentation License	1814
1. APPLICABILITY AND DEFINITIONS	1815
2. VERBATIM COPYING	1818
3. COPYING IN QUANTITY	1818
4. MODIFICATIONS	1819
5. COMBINING DOCUMENTS	1823
6. COLLECTIONS OF DOCUMENTS	1823
7. AGGREGATION WITH INDEPENDENT WORKS	1824
8. TRANSLATION	1824
9. TERMINATION	1825
10. FUTURE REVISIONS OF THIS LICENSE	1825
ADDENDUM: How to use this License for your documents	1826

Part T Topics

F Fields	1829
F Fields	1830
FF Finite Fields	1833
EXC Exercises	1841
SOL Solutions	1844
T Trace	1845
EXC Exercises	1853

SOL Solutions	1855
HP Hadamard Product	1857
DMHP Diagonal Matrices and the Hadamard Product	1863
EXC Exercises	1868
VM Vandermonde Matrix	1870
PSM Positive Semi-definite Matrices	1878
PSM Positive Semi-Definite Matrices	1879
EXC Exercises	1884
Chapter MD Matrix Decompositions	1885
ROD Rank One Decomposition	1887
TD Triangular Decomposition	1897
TD Triangular Decomposition	1897
TDSSE Triangular Decomposition and Solving Systems of Equations \ldots	1904
CTD Computing Triangular Decompositions	1907
SVD Singular Value Decomposition	1914
MAP Matrix-Adjoint Product	1915
SVD Singular Value Decomposition	1923
SR Square Roots	1926
SRM Square Root of a Matrix	1927
POD Polar Decomposition	1935

Part A Applications

CF Curve Fitting	1940
DF Data Fitting	1942
EXC Exercises	1948
SAS Sharing A Secret	1949

Contributors

Beezer, David. Belarmine Preparatory School, Tacoma Beezer, Robert. University of Puget Sound http://buzzard.ups.edu/

Black, Chris.

Braithwaite, David. Chicago, Illinois

Bucht, Sara. University of Puget Sound

Canfield, Steve. University of Puget Sound

Hubert, Dupont. Créteil, France

Fellez, Sarah. University of Puget Sound

Fickenscher, Eric. University of Puget Sound

Jackson, Martin. University of Puget Sound http://www.math.ups.edu/~martinj

Kessler, Ivan. University of Puget Sound

- Kreher, Don. Michigan Technological University http://www.math.mtu.edu/~kreher/
- Hamrick, Mark. St. Louis University
- Linenthal, Jacob. University of Puget Sound
- Million, Elizabeth. University of Puget Sound
- Osborne, Travis. University of Puget Sound
- Riegsecker, Joe. Middlebury, Indiana joepye (at) pobox (dot) com
- Perkel, Manley. University of Puget Sound
- Phelps, Douglas. University of Puget Sound
- Shoemaker, Mark. University of Puget Sound
- Toth, Zoltan. http://zoli.web.elte.hu
- Zimmer, Andy. University of Puget Sound

Definitions

Section	WILA	
Section	SSLE	
SLE	System of Linear Equations	18
SSLE	Solution of a System of Linear Equations	18
SSSLE	Solution Set of a System of Linear Equations	19
ESYS	Equivalent Systems	22
EO	Equation Operations	23
Section	RREF	
Μ	Matrix	50
\mathbf{CV}	Column Vector	52
ZCV	Zero Column Vector	53

CM VOC SOLV MRLS AM RO REM RREF RR	Vector of Constants5Solution Vector5Matrix Representation of a Linear System5Augmented Matrix5Row Operations5Row-Equivalent Matrices6Reduced Row-Echelon Form6	53 54 55 56 57 59 60 63 85
Section TS CS IDV	S Consistent System	13 18
Section HS HS TSHSE NSM	Homogeneous System14Trivial Solution to Homogeneous Systems of Equations14	46 48 52
Section NM SQM NM IM	Square Matrix17Nonsingular Matrix17	73 73 75

Section VC	
VSCV	Vector Space of Column Vectors
CVE	Column Vector Equality
CVA	Column Vector Addition
CVSM	Column Vector Scalar Multiplication
Section LC	
LCCV	Linear Combination of Column Vectors 221
Section SS	
SSCV	Span of a Set of Column Vectors 270
Section LI	
RLDCV	Relation of Linear Dependence for Column Vectors316316
LICV	Linear Independence of Column Vectors 317
Section LD	S
Section O	
CCCV	Complex Conjugate of a Column Vector
IP	Inner Product
NV OV	Norm of a Vector 404 Orthogonal Vector 408
OV	Orthogonal Vectors

OSV	Orthogonal Set of Vectors	09
SUV	Standard Unit Vectors	09
ONS	OrthoNormal Set	18
Section M	C	
VSM	Vector Space of $m \times n$ Matrices	26
ME	Matrix Equality 44	27
MA		27
MSM	Matrix Scalar Multiplication	28
ZM	Zero Matrix	33
TM		33
SYM	Symmetric Matrix	34
CCM	Complex Conjugate of a Matrix	38
А		42
Section MI	M	
MVP	Matrix-Vector Product	61
MM	Matrix Multiplication	68
HM	Hermitian Matrix	86
Section Ml	ISLE	
MI	Matrix Inverse	06

Section MINM			
UM	Unitary Matrices	44	
Section CF	\mathbf{s}		
CSM	Column Space of a Matrix	60	
RSM	Row Space of a Matrix	77	
Section FS			
LNS	The second se	09	
EEF	Extended Echelon Form	19	
Section VS			
VS	Vector Space	59	
Section S			
S	Subspace	94	
TS	Trivial Subspaces	04	
LC		07	
SS	Span of a Set	09	
Section LIS	SS		
RLD	Relation of Linear Dependence77	35	
LI	Linear Independence 7	35	

TSVS	To Span a Vector Space 7	46
Section B B	Basis	77
Section D D NOM ROM	Nullity Of a Matrix 8	15 29 29
Section PD DS		64
Section DN	Л	
ELEM		83
\mathbf{SM}		93
DM	Determinant of a Matrix	94
Section PD Section EE		
EEM	Eigenvalues and Eigenvectors of a Matrix	45
CP		60
EM	Eigenspace of a Matrix	63

AME	Algebraic Multiplicity of an Eigenvalue	8
GME	Geometric Multiplicity of an Eigenvalue	9
Section PE	E	
Section SD		
SIM	Similar Matrices	9
DIM	Diagonal Matrix	7
DZM	Diagonalizable Matrix	7
Section LT		
LT	Linear Transformation107	2
PI	Pre-Image 110	1
LTA	Linear Transformation Addition	5
LTSM	Linear Transformation Scalar Multiplication	8
LTC	Linear Transformation Composition	1
Section ILT		
ILT	Injective Linear Transformation	8
KLT	Kernel of a Linear Transformation	8
Section SL	Г	
SLT	Surjective Linear Transformation	9
RLT	Range of a Linear Transformation 117	8

'LT		
Identity Linear Transformation	1213	
Invertible Linear Transformations	1214	
Isomorphic Vector Spaces	1230	
Rank Of a Linear Transformation	1233	
Nullity Of a Linear Transformation	1234	
R		
Vector Representation	1261	
Section MR		
Matrix Representation	1286	
В		
Eigenvalue and Eigenvector of a Linear Transformation	1356	
Change-of-Basis Matrix	1359	
D		
Upper Triangular Matrix	1415	
Lower Triangular Matrix	1415	
Normal Matrix	1427	
E	Identity Linear Transformation Invertible Linear Transformations Isomorphic Vector Spaces Invertible Linear Transformation Rank Of a Linear Transformation Invertible Linear Transformation Nullity Of a Linear Transformation Invertible Linear Transformation Nullity Of a Linear Transformation Invertible Linear Transformation R Matrix Representation R Matrix Representation S Eigenvalue and Eigenvector of a Linear Transformation Change-of-Basis Matrix Invertible Linear Transformation D Upper Triangular Matrix Lower Triangular Matrix Invertible Linear Transformation	

Section NL	Т
NLT	Nilpotent Linear Transformation
JB	Jordan Block
Section IS	
IS	Invariant Subspace
GEV	Generalized Eigenvector
GES	Generalized Eigenspace
LTR	Linear Transformation Restriction
IE	Index of an Eigenvalue 1508
Section JC.	F
JCF	Jordan Canonical Form
Section CN	0
CNE	Complex Number Equality
CNA	Complex Number Addition
CNM	Complex Number Multiplication
CCN	Conjugate of a Complex Number
MCN	Modulus of a Complex Number
Section SE	Г
SET	Set 1602

SSET	Subset)3
ES	Empty Set)3
SE	Set Equality)4
С	Cardinality)6
SU	Set Union)7
SI	Set Intersection)8
\mathbf{SC}	Set Complement)8
Section PT		
Section F		
F	Field	30
IMP	Integers Modulo a Prime	33
Section T		
Т	Trace	15
Section HF		
HP	Hadamard Product	57
HID	Hadamard Identity	59
HI	Hadamard Inverse	30
Section VN	1	
VM	Vandermonde Matrix	70

Section PSM	1070
PSM Positive Semi-Definite Matrix	1879
Section ROD	
Section TD	
Section SVD	
SV Singular Values	1923
Section SR	
SRM Square Root of a Matrix	1934
Section POD	
Section CF	
LSS Least Squares Solution	1944
Section SAS	

Theorems

Section WI Section SS		
EOPSS	Equation Operations Preserve Solution Sets	24
Section RF	REF	
REMES	Row-Equivalent Matrices represent Equivalent Systems	61
REMEF	Row-Equivalent Matrix in Echelon Form	66
RREFU	Reduced Row-Echelon Form is Unique	70
Section TS	S	
RCLS	Recognizing Consistency of a Linear System	120
ISRN	Inconsistent Systems, r and n	122

CSRN	Consistent Systems, r and n	123
FVCS	Free Variables for Consistent Systems	124
PSSLS	Possible Solution Sets for Linear Systems	126
CMVEI	Consistent, More Variables than Equations, Infinite solutions	127
Section HS	SE	
HSC	Homogeneous Systems are Consistent	147
HMVEI	Homogeneous, More Variables than Equations, Infinite solutions	151
Section NN	M	
NMRRI	Nonsingular Matrices Row Reduce to the Identity matrix	176
NMTNS	Nonsingular Matrices have Trivial Null Spaces	179
NMUS	Nonsingular Matrices and Unique Solutions	180
NME1	Nonsingular Matrix Equivalences, Round 1	181
Section VO)	
VSPCV	Vector Space Properties of Column Vectors	205
Section LC	<u> </u>	
SLSLC	Solutions to Linear Systems are Linear Combinations	229
VFSLS	Vector Form of Solutions to Linear Systems	244
PSPHS	Particular Solution Plus Homogeneous Solutions	256

Section SS	
SSNS	Spanning Sets for Null Spaces 283
Section LI	
LIVHS	Linearly Independent Vectors and Homogeneous Systems
LIVRN	Linearly Independent Vectors, r and n
MVSLD	More Vectors than Size implies Linear Dependence
NMLIC	Nonsingular Matrices have Linearly Independent Columns
NME2	Nonsingular Matrix Equivalences, Round 2
BNS	Basis for Null Spaces 334
Section LD	S
DLDS	Dependency in Linearly Dependent Sets
BS	Basis of a Span
Section O	
CRVA	Conjugation Respects Vector Addition
CRSM	Conjugation Respects Vector Scalar Multiplication
IPVA	Inner Product and Vector Addition
IPSM	Inner Product and Scalar Multiplication
IPAC	Inner Product is Anti-Commutative
IPN	Inner Products and Norms
PIP	Positive Inner Products406406

OSLI	Orthogonal Sets are Linearly Independent	412
GSP	Gram-Schmidt Procedure	413
Section M	0	
VSPM	Vector Space Properties of Matrices	430
SMS	Symmetric Matrices are Square	435
TMA	Transpose and Matrix Addition	436
TMSM	Transpose and Matrix Scalar Multiplication	436
TT	Transpose of a Transpose	437
CRMA	Conjugation Respects Matrix Addition	439
CRMSM	Conjugation Respects Matrix Scalar Multiplication	440
CCM	Conjugate of the Conjugate of a Matrix	440
MCT	Matrix Conjugation and Transposes	441
AMA	Adjoint and Matrix Addition	442
AMSM	Adjoint and Matrix Scalar Multiplication	443
AA	Adjoint of an Adjoint	443
Section MI	М	
SLEMM	Systems of Linear Equations as Matrix Multiplication	462

SLEMM	Systems of Linear Equations as Matrix Multiplication	462
EMMVP	Equal Matrices and Matrix-Vector Products	466
EMP	Entries of Matrix Products	471
MMZM	Matrix Multiplication and the Zero Matrix	474
MMIM	Matrix Multiplication and Identity Matrix	475

MMDAA	Matrix Multiplication Distributes Across Addition	476
MMSMM	Matrix Multiplication and Scalar Matrix Multiplication	477
MMA	Matrix Multiplication is Associative	478
MMIP	Matrix Multiplication and Inner Products	480
MMCC	Matrix Multiplication and Complex Conjugation	481
MMT	Matrix Multiplication and Transposes	482
MMAD	Matrix Multiplication and Adjoints	483
AIP	Adjoint and Inner Product	485
HMIP	Hermitian Matrices and Inner Products	486

Section MISLE

TTMI	Two-by-Two Matrix Inverse	509
CINM	Computing the Inverse of a Nonsingular Matrix	516
MIU	Matrix Inverse is Unique	519
SS	Socks and Shoes	520
MIMI	Matrix Inverse of a Matrix Inverse	521
MIT	Matrix Inverse of a Transpose	522
MISM	Matrix Inverse of a Scalar Multiple	523

Section MINM

NPNT	Nonsingular Product has Nonsingular Terms	538
OSIS	One-Sided Inverse is Sufficient	541
NI	Nonsingularity is Invertibility	542

SNCM Solution with Nonsingular Coefficient Matrix	r 49
	543
UMI Unitary Matrices are Invertible	546
CUMOS Columns of Unitary Matrices are Orthonormal Sets	547
UMPIP Unitary Matrices Preserve Inner Products	549
Section CRS	
CSCS Column Spaces and Consistent Systems	563
BCS Basis of the Column Space	569
CSNM Column Space of a Nonsingular Matrix	574
NME4 Nonsingular Matrix Equivalences, Round 4	575
REMRS Row-Equivalent Matrices have equal Row Spaces	580
BRS Basis for the Row Space	583
CSRST Column Space, Row Space, Transpose	586
Section FS	
PEEF Properties of Extended Echelon Form	621
FS Four Subsets	625
Section VS	
ZVU Zero Vector is Unique	675
AIU Additive Inverses are Unique	675
ZSSM Zero Scalar in Scalar Multiplication	676

ZVSM	Zero Vector in Scalar Multiplication	677
AISM	Additive Inverses from Scalar Multiplication	678
SMEZV	Scalar Multiplication Equals the Zero Vector	679
Section S		
TSS	Testing Subsets for Subspaces	698
NSMS	Null Space of a Matrix is a Subspace	705
SSS	Span of a Set is a Subspace	709
CSMS	Column Space of a Matrix is a Subspace	719
RSMS	Row Space of a Matrix is a Subspace	719
LNSMS	Left Null Space of a Matrix is a Subspace	720
Section LIS	SS	
VRRB	Vector Representation Relative to a Basis	755
Section B		
SUVB	Standard Unit Vectors are a Basis	779
CNMB	Columns of Nonsingular Matrix are a Basis	789
NME5	Nonsingular Matrix Equivalences, Round 5	791
COB	Coordinates and Orthonormal Bases	793
UMCOB	Unitary Matrices Convert Orthonormal Bases	798

Section D

SSLD	Spanning Sets and Linear Dependence	816
BIS	Bases have Identical Sizes	823
DCM	Dimension of \mathbb{C}^m	824
DP	Dimension of P_n	824
DM	Dimension of M_{mn}	824
CRN	Computing Rank and Nullity	830
RPNC	Rank Plus Nullity is Columns	831
RNNM	Rank and Nullity of a Nonsingular Matrix	833
NME6	Nonsingular Matrix Equivalences, Round 6	834
Section P	D	
ELIS	Extending Linearly Independent Sets	850
G	Goldilocks	852
PSSD	Proper Subspaces have Smaller Dimension	857
EDYES	Equal Dimensions Yields Equal Subspaces	858
RMRT	Rank of a Matrix is the Rank of the Transpose	859
DFS	Dimensions of Four Subspaces	861
DSFB	Direct Sum From a Basis	865
DSFOS	Direct Sum From One Subspace	867
DSZV	Direct Sums and Zero Vectors	868
DSZI	Direct Sums and Zero Intersection	869
DSLI	Direct Sums and Linear Independence	870
DSD	Direct Sums and Dimension	872

RDS	Repeated Direct Sums	873
Section DI	М	
EMDRO	Elementary Matrices Do Row Operations	886
EMN	Elementary Matrices are Nonsingular	891
NMPEM	Nonsingular Matrices are Products of Elementary Matrices	892
DMST	Determinant of Matrices of Size Two	896
DER	Determinant Expansion about Rows	897
DT	Determinant of the Transpose	899
DEC	Determinant Expansion about Columns	901
Section PI	DM	
DZRC	Determinant with Zero Row or Column	919
DRCS	Determinant for Row or Column Swap	920
DRCM	Determinant for Row or Column Multiples	922
DERC	Determinant with Equal Rows or Columns	923
DRCMA	Determinant for Row or Column Multiples and Addition	924
DIM	Determinant of the Identity Matrix	930
DEM	Determinants of Elementary Matrices	932
DEMMM	Determinants, Elementary Matrices, Matrix Multiplication	933
SMZD	Singular Matrices have Zero Determinants	935
NME7	Nonsingular Matrix Equivalences, Round 7	936
DRMM	Determinant Respects Matrix Multiplication	938

Section EE		
EMHE	Every Matrix Has an Eigenvalue	953
EMRCP	Eigenvalues of a Matrix are Roots of Characteristic Polynomials	962
EMS	Eigenspace for a Matrix is a Subspace	963
EMNS	Eigenspace of a Matrix is a Null Space	965
Section PE	CE	
EDELI	Eigenvectors with Distinct Eigenvalues are Linearly Independent	1001
SMZE	Singular Matrices have Zero Eigenvalues	1003
NME8	Nonsingular Matrix Equivalences, Round 8	1004
ESMM	Eigenvalues of a Scalar Multiple of a Matrix	1005
EOMP	Eigenvalues Of Matrix Powers	1006
EPM	Eigenvalues of the Polynomial of a Matrix	1007
EIM	Eigenvalues of the Inverse of a Matrix	1009
ETM	Eigenvalues of the Transpose of a Matrix	1011
ERMCP	Eigenvalues of Real Matrices come in Conjugate Pairs	1012
DCP	Degree of the Characteristic Polynomial	1013
NEM	Number of Eigenvalues of a Matrix	1015
ME	Multiplicities of an Eigenvalue	1016
MNEM	Maximum Number of Eigenvalues of a Matrix	1019
HMRE	Hermitian Matrices have Real Eigenvalues	1020
HMOE	Hermitian Matrices have Orthogonal Eigenvectors	1022

Section SD	
SER	Similarity is an Equivalence Relation
SMEE	Similar Matrices have Equal Eigenvalues
DC	Diagonalization Characterization
DMFE	Diagonalizable Matrices have Full Eigenspaces 1044
DED	Distinct Eigenvalues implies Diagonalizable
Section LT	
LTTZZ	Linear Transformations Take Zero to Zero
MBLT	Matrices Build Linear Transformations
MLTCV	Matrix of a Linear Transformation, Column Vectors
LTLC	Linear Transformations and Linear Combinations
LTDB	Linear Transformation Defined on a Basis
SLTLT	Sum of Linear Transformations is a Linear Transformation
MLTLT	Multiple of a Linear Transformation is a Linear Transformation 1108
VSLT	Vector Space of Linear Transformations
CLTLT	Composition of Linear Transformations is a Linear Transformation 1111

Section ILT

KLTS	Kernel of a Linear Transformation is a Subspace	1140
KPI	Kernel and Pre-Image	1143
KILT	Kernel of an Injective Linear Transformation	1144

ILTLI	Injective Linear Transformations and Linear Independence	1148
ILTB	Injective Linear Transformations and Bases	1149
ILTD	Injective Linear Transformations and Dimension	1150
CILTI	Composition of Injective Linear Transformations is Injective	1152
Section SL	Л	
RLTS	Range of a Linear Transformation is a Subspace	1181
RSLT	Range of a Surjective Linear Transformation	1185
SSRLT	Spanning Set for Range of a Linear Transformation	1188
RPI	Range and Pre-Image	1191
SLTB	Surjective Linear Transformations and Bases	1192
SLTD	Surjective Linear Transformations and Dimension	1193
CSLTS	Composition of Surjective Linear Transformations is Surjective	1195

Section IVLT

ILTLT	Inverse of a Linear Transformation is a Linear Transformation	1219
IILT	Inverse of an Invertible Linear Transformation	1220
ILTIS	Invertible Linear Transformations are Injective and Surjective	1221
CIVLT	Composition of Invertible Linear Transformations	1227
ICLT	Inverse of a Composition of Linear Transformations	1227
IVSED	Isomorphic Vector Spaces have Equal Dimension	1232
ROSLT	Rank Of a Surjective Linear Transformation	1234
NOILT	Nullity Of an Injective Linear Transformation	1235

RPNDD	Rank Plus Nullity is Domain Dimension	1235
Section V	R	
VRLT	Vector Representation is a Linear Transformation	1262
VRI	Vector Representation is Injective	1270
VRS	Vector Representation is Surjective	1271
VRILT	Vector Representation is an Invertible Linear Transformation	1272
CFDVS	Characterization of Finite Dimensional Vector Spaces	1272
IFDVS	Isomorphism of Finite Dimensional Vector Spaces	1274
CLI	Coordinatization and Linear Independence	1275
CSS	Coordinatization and Spanning Sets	1276
Section M	R	
FTMR	Fundamental Theorem of Matrix Representation	1293
MRSLT	Matrix Representation of a Sum of Linear Transformations	1300
MRMLT	Matrix Representation of a Multiple of a Linear Transformation	1301
MRCLT	Matrix Representation of a Composition of Linear Transformations	1302
KNSI	Kernel and Null Space Isomorphism	1311
RCSI	Range and Column Space Isomorphism	1317
IMR	Invertible Matrix Representations	1322

Invertible Matrices, Invertible Linear Transformation

IMILT

NME9

1327

1328

3	
Change-of-Basis	1359
Inverse of Change-of-Basis Matrix	1360
Matrix Representation and Change of Basis	1371
Similarity and Change of Basis	1377
Eigenvalues, Eigenvectors, Representations	1384
)	
Product of Triangular Matrices is Triangular	1415
Inverse of a Triangular Matrix is Triangular	1417
Upper Triangular Matrix Representation	1419
Orthonormal Basis for Upper Triangular Representation	1424
Orthonormal Diagonalization	1429
Orthonormal Bases and Normal Matrices	1434
T	
Nilpotent Jordan Blocks	1447
Eigenvalues of Nilpotent Linear Transformations	1448
Diagonalizable Nilpotent Linear Transformations	1449
Kernels of Powers of Linear Transformations	1450
Kernels of Powers of Nilpotent Linear Transformations	1453
Canonical Form for Nilpotent Linear Transformations	1458
	Change-of-Basis

Section IS	
EIS	Eigenspaces are Invariant Subspaces
KPIS	Kernels of Powers are Invariant Subspaces
GESIS	Generalized Eigenspace is an Invariant Subspace
GEK	Generalized Eigenspace as a Kernel 1487
RGEN	Restriction to Generalized Eigenspace is Nilpotent 1507
MRRGE	Matrix Representation of a Restriction to a Generalized Eigenspace 1514
Section JC	F
GESD	Generalized Eigenspace Decomposition
DGES	Dimension of Generalized Eigenspaces
JCFLT	Jordan Canonical Form for a Linear Transformation
CHT	Cayley-Hamilton Theorem 1563
Section CN	10
PCNA	Properties of Complex Number Arithmetic
CCRA	Complex Conjugation Respects Addition
CCRM	Complex Conjugation Respects Multiplication
CCT	Complex Conjugation Twice
Section SE	Т

Section PT

Section F

FIMP	Field of Integers Modulo a Prime	1833
Section T		
TL	Trace is Linear	1846
TSRM	Trace is Symmetric with Respect to Multiplication	1847
TIST	Trace is Invariant Under Similarity Transformations	1848
TSE	Trace is the Sum of the Eigenvalues	1849
Section HI	D	
HPC	Hadamard Product is Commutative	1858
HPHID	Hadamard Product with the Hadamard Identity	1859
HPHI	Hadamard Product with Hadamard Inverses	1860
HPDAA	Hadamard Product Distributes Across Addition	1861
HPSMM	Hadamard Product and Scalar Matrix Multiplication	1862
DMHP	Diagonalizable Matrices and the Hadamard Product	1863
DMMP	Diagonal Matrices and Matrix Products	1865
Section VN	M	
DVM	Determinant of a Vandermonde Matrix	1871
NVM	Nonsingular Vandermonde Matrix	1877
Section PS	SM	
CPSM	Creating Positive Semi-Definite Matrices	1879

EPSM	Eigenvalues of Positive Semi-definite Matrices	1880
Section RC ROD	D Rank One Decomposition	1887
Section TI TD TDEE) Triangular Decomposition	1898 1908
Section SV EEMAP SVD	D Eigenvalues and Eigenvectors of Matrix-Adjoint Product Singular Value Decomposition	1915 1924
Section SR PSMSR EESR USR	Positive Semi-Definite Matrices and Square Roots Eigenvalues and Eigenspaces of a Square Root Unique Square Root	1927 1930 1933
Section PC PDM	DD Polar Decomposition of a Matrix	1936
Section CF IP	Interpolating Polynomial	1940

LSMR	Least Squares Minimizes Residuals	1944
------	-----------------------------------	------

Section SAS

Notation

Μ	$A: Matrix \ldots \ldots$	51
MC	$[A]_{ij}$: Matrix Components	51
CV	\mathbf{v} : Čolumn Vector	52
CVC	$[\mathbf{v}]_i$: Column Vector Components	52
ZCV	0 : Zero Column Vector	53
MRLS	$\mathcal{LS}(A, \mathbf{b})$: Matrix Representation of a Linear System	56
AM	$[A \mid \mathbf{b}]$: Augmented Matrix	57
RO	$R_i \leftrightarrow R_j, \alpha R_i, \alpha R_i + R_j$: Row Operations	60
RREFA	r, D, F: Reduced Row-Echelon Form Analysis	64
NSM	$\mathcal{N}(A)$: Null Space of a Matrix	52
IM	I_m : Identity Matrix	75
VSCV	\mathbb{C}^m : Vector Space of Column Vectors	98

CVE	$\mathbf{u} = \mathbf{v}$: Column Vector Equality	199
CVA	$\mathbf{u} + \mathbf{v}$: Column Vector Addition	201
CVSM	$\alpha \mathbf{u}$: Column Vector Scalar Multiplication	203
SSV	$\langle S \rangle$: Span of a Set of Vectors	271
CCCV	$\overline{\mathbf{u}}$: Complex Conjugate of a Column Vector	396
IP	$\langle \mathbf{u}, \mathbf{v} \rangle$: Inner Product	398
NV	$\ \mathbf{v}\ $: Norm of a Vector	404
SUV	\mathbf{e}_i : Standard Unit Vectors	410
VSM	M_{mn} : Vector Space of Matrices	426
ME	A = B: Matrix Equality	427
MA	A + B: Matrix Addition	427
MSM	αA : Matrix Scalar Multiplication	428
ZM	\mathcal{O} : Zero Matrix	433
TM	A^t : Transpose of a Matrix	433
CCM	\overline{A} : Complex Conjugate of a Matrix	438
А	$A^*: \operatorname{Adjoint} \ldots \ldots$	442
MVP	Au: Matrix-Vector Product \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots	461
MI	A^{-1} : Matrix Inverse	506
CSM	$\mathcal{C}(A)$: Column Space of a Matrix $\ldots \ldots \ldots$	560
RSM	$\mathcal{R}(A)$: Row Space of a Matrix	577
LNS	$\mathcal{L}(A)$: Left Null Space	609
D	$\dim(V)$: Dimension	815
NOM	n(A): Nullity of a Matrix	829

ROM	r(A): Rank of a Matrix	829
DS	$V = U \oplus W$: Direct Sum	864
ELEM	$E_{i,j}, E_i(\alpha), E_{i,j}(\alpha)$: Elementary Matrix	885
SM		893
DM		894
AME	$\alpha_A(\lambda)$: Algebraic Multiplicity of an Eigenvalue	968
GME	$\gamma_A(\lambda)$: Geometric Multiplicity of an Eigenvalue	969
LT	$T: U \to V$: Linear Transformation	.073
KLT	$\mathcal{K}(T)$: Kernel of a Linear Transformation	138
RLT	$\mathcal{R}(T)$: Range of a Linear Transformation	178
ROLT	r(T): Rank of a Linear Transformation	234
NOLT	n(T): Nullity of a Linear Transformation	234
VR	$\rho_B(\mathbf{w})$: Vector Representation	261
MR	$M_{B,C}^T$: Matrix Representation	286
JB		441
GES	$\mathcal{G}_T(\lambda)$: Generalized Eigenspace	485
LTR		495
IE		508
CNE	$\alpha = \beta$: Complex Number Equality	596
CNA	$\alpha + \beta$: Complex Number Addition	596
CNM	$\alpha\beta$: Complex Number Multiplication	596
CCN		598
SETM		602

SSET	$S \subseteq T$: Subset	1603
ES	\emptyset : Empty Set	1603
SE	S = T: Set Equality	1604
С	S : Cardinality	1606
SU	$S \cup T$: Set Union	1607
SI	$S \cap T$: Set Intersection	1608
\mathbf{SC}	\overline{S} : Set Complement	1609
Т	t(A): Trace	1845
HP	$A \circ B$: Hadamard Product	1857
HID	J_{mn} : Hadamard Identity	1859
HI	\widehat{A} : Hadamard Inverse	1860
SRM	$A^{1/2}$: Square Root of a Matrix	1934

Diagrams

DTSLS	Decision Tree for Solving Linear Systems	126
CSRST	Column Space and Row Space Techniques	643
DLTA	Definition of Linear Transformation, Additive	1073
DLTM	Definition of Linear Transformation, Multiplicative	1074
GLT	General Linear Transformation	1082
NILT	Non-Injective Linear Transformation	1131
ILT	Injective Linear Transformation	1135
FTMR	Fundamental Theorem of Matrix Representations	1295
FTMRA	Fundamental Theorem of Matrix Representations (Alternate)	1295
MRCLT	Matrix Representation and Composition of Linear Transformations	1310

Examples

Section	WILA	
TMP	Trail Mix Packaging	6
Section	SSLE	
STNE	Solving two (nonlinear) equations	16
NSE	Notation for a system of equations	19
TTS	Three typical systems	21
US	Three equations, one solution	29
IS	Three equations, infinitely many solutions	31
Section	RREF	
AM	A matrix	51

NSLE	Notation for systems of linear equations	56
AMAA	Augmented matrix for Archetype A	58
TREM	Two row-equivalent matrices	60
USR	Three equations, one solution, reprised	62
RREF	A matrix in reduced row-echelon form	65
NRREF	A matrix not in reduced row-echelon form	65
SAB	Solutions for Archetype B	79
SAA	Solutions for Archetype A	81
SAE	Solutions for Archetype E	83

Section TSS

RREFN	Reduced row-echelon form notation	113
ISSI	Describing infinite solution sets, Archetype I	115
FDV	Free and dependent variables	118
CFV	Counting free variables	124
OSGMD	One solution gives many, Archetype D	127

Section HSE

AHSAC	Archetype C as a homogeneous system	147
HUSAB	Homogeneous, unique solution, Archetype B	148
HISAA	Homogeneous, infinite solutions, Archetype A	149
HISAD	Homogeneous, infinite solutions, Archetype D	150
NSEAI	Null space elements of Archetype I	153

CNS2	Computing a null space, $\#2$	
		56
Section NN	М	
		74
S		74
NM	\mathbf{O}	74
IM	An identity matrix	75
SRR	Singular matrix, row-reduced	76
NSR	Nonsingular matrix, row-reduced	77
NSS		78
NSNM		79
Section VO	C	
VESE	Vector equality for a system of equations	00
VA)2
CVSM)4
U V SIVI	Scalar multiplication in \mathbb{C}	J4
Section LO	C	
		22
TLC	Two linear combinations in \mathbb{C}^6	22 25
TLC ABLC	Two linear combinations in \mathbb{C}^6 22Archetype B as a linear combination22	25
TLC ABLC AALC	Two linear combinations in \mathbb{C}^6 22Archetype B as a linear combination22Archetype A as a linear combination2222	25 27
TLC ABLC	Two linear combinations in \mathbb{C}^6 22Archetype B as a linear combination22Archetype A as a linear combination22Vector form of solutions for Archetype D23	25

VFSAI	Vector form of solutions for Archetype I
VFSAL	Vector form of solutions for Archetype L
PSHS	Particular solutions, homogeneous solutions, Archetype D
Section SS	
ABS	A basic span
SCAA	Span of the columns of Archetype A
SCAB	Span of the columns of Archetype B
SSNS	Spanning set of a null space
NSDS	Null space directly as a span 287
SCAD	Span of the columns of Archetype D
Section LI	
TDC	
LDS	Linearly dependent set in \mathbb{C}^5
LDS LIS	
LIS	Linearly independent set in \mathbb{C}^5 320Linearly independent, homogeneous system322
LIS LIHS	Linearly independent set in \mathbb{C}^5 320Linearly independent, homogeneous system322
LIS LIHS LDHS	Linearly independent set in \mathbb{C}^5 320Linearly independent, homogeneous system322Linearly dependent, homogeneous system323
LIS LIHS LDHS LDRN	Linearly independent set in \mathbb{C}^5 320Linearly independent, homogeneous system322Linearly dependent, homogeneous system323Linearly dependent, $r < n$ 326326
LIS LIHS LDHS LDRN LLDS	Linearly independent set in \mathbb{C}^5 320Linearly independent, homogeneous system322Linearly dependent, homogeneous system323Linearly dependent, $r < n$ 326Large linearly dependent set in \mathbb{C}^4 327
LIS LIHS LDHS LDRN LLDS LDCAA	Linearly independent set in \mathbb{C}^5 320Linearly independent, homogeneous system322Linearly dependent, homogeneous system323Linearly dependent, $r < n$ 326Large linearly dependent set in \mathbb{C}^4 326Linearly dependent columns in Archetype A328

Section LD	ho S
RSC5	Reducing a span in \mathbb{C}^5
COV	Casting out vectors
RSC4	Reducing a span in \mathbb{C}^4
RES	Reworking elements of a span
Section O	
CSIP	Computing some inner products
CNSV	Computing the norm of some vectors
TOV	Two orthogonal vectors $\ldots \ldots \ldots$
SUVOS	Standard Unit Vectors are an Orthogonal Set
AOS	An orthogonal set $\ldots \ldots 410$
GSTV	Gram-Schmidt of three vectors
ONTV	Orthonormal set, three vectors
ONFV	Orthonormal set, four vectors
Section MO)
MA	Addition of two matrices in M_{23}
MSM	Scalar multiplication in M_{32}
TM	Transpose of a 3×4 matrix
SYM	A symmetric 5×5 matrix
CCM	Complex conjugate of a matrix

Section MM

MTV	A matrix times a vector	461
MNSLE	Matrix notation for systems of linear equations	463
MBC	Money's best cities	464
PTM	Product of two matrices	468
MMNC	Matrix multiplication is not commutative	470
PTMEE	Product of two matrices, entry-by-entry	472
Section MI	ISLE	
SABMI	Solutions to Archetype B with a matrix inverse	503
MWIAA	A matrix without an inverse, Archetype A	506
MI	Matrix inverse	507
CMI	Computing a matrix inverse	512
CMIAB	Computing a matrix inverse, Archetype B	517
Section MI	INM	
UM3	Unitary matrix of size 3	545
UPM	Unitary permutation matrix	545
OSMC	Orthonormal set from matrix columns	548
Section CF	RS	
CSMCS	Column space of a matrix and consistent systems	561

MCSM	Membership in the column space of a matrix	564
CSTW	Column space, two ways	568
CSOCD	Column space, original columns, Archetype D	571
CSAA	Column space of Archetype A	573
CSAB	Column space of Archetype B	574
RSAI	Row space of Archetype I	577
RSREM	Row spaces of two row-equivalent matrices	582
IAS	Improving a span	585
CSROI	Column space from row operations, Archetype I	587
Section FS		610
CSANS	1	612
SEEF	1 1	619
FS1		634
FS2		635
FSAG	Four subsets, Archetype G	638
Section VS	5	
VSCV	The vector space \mathbb{C}^m	663
VSM	The vector space of matrices, M_{mn}	663
VSP	The vector space of polynomials, P_n	665
VSIS	The vector space of infinite sequences	667

VSF	The vector space of functions	67
VSS	The singleton vector space	68
CVS	The crazy vector space	70
PCVS	Properties for the Crazy Vector Space	79
Section S		
SC3	A subspace of \mathbb{C}^3	94
SP4		00
NSC2Z		02
NSC2A	A non-subspace in \mathbb{C}^2 , additive closure	03
NSC2S	A non-subspace in \mathbb{C}^2 , scalar multiplication closure	03
RSNS	Recasting a subspace as a null space	06
LCM	A linear combination of matrices	07
SSP	Span of a set of polynomials	12
SM32	A subspace of M_{32}	'14
Section LI	SS	
LIP4	Linear independence in P_4	36
LIM32	Linear independence in M_{32}	39
LIC	Linearly independent set in the crazy vector space	43
SSP4	Spanning set in P_4	46
SSM22	Spanning set in M_{22}	49
SSC	Spanning set in the crazy vector space	51

EXAMPLES lxxii

AVR	A vector representation	54
Section B		
BP	Bases for P_n	30
BM	A basis for the vector space of matrices	30
BSP4	A basis for a subspace of P_4	31
BSM22	A basis for a subspace of M_{22}	32
BC	Basis for the crazy vector space	34
RSB	Row space basis	35
RS	Reducing a span	37
CABAK	Columns as Basis, Archetype K	90
CROB4	Coordinatization relative to an orthonormal basis, \mathbb{C}^4	94
CROB3	Coordinatization relative to an orthonormal basis, \mathbb{C}^3	96
Section D		
LDP4	Linearly dependent set in P_4	22
DSM22	Dimension of a subspace of M_{22}	25
DSP4	Dimension of a subspace of P_4	27
DC	Dimension of the crazy vector space	27
VSPUD	Vector space of polynomials with unbounded degree	28
RNM	Rank and nullity of a matrix8282	29
RNSM	Rank and nullity of a square matrix83	32

Section PD		
BPR	Bases for P_n , reprised	354
BDM22	Basis by dimension in M_{22}	354
SVP4	Sets of vectors in P_4	356
RRTI	Rank, rank of transpose, Archetype I 8	860
SDS	Simple direct sum 8	865
Section DM	Ν	
EMRO	Elementary matrices and row operations	885
\mathbf{SS}	Some submatrices	393
D33M	Determinant of a 3×3 matrix $\ldots \ldots \ldots$	395
TCSD	Two computations, same determinant	002
DUTM	Determinant of an upper triangular matrix	903
Section PD	DM	
DRO	Determinant by row operations	926
ZNDAB	Zero and nonzero determinant, Archetypes A and B	936
Section EE		
SEE	Some eigenvalues and eigenvectors	946
\mathbf{PM}	Polynomial of a matrix	949
CAEHW	Computing an eigenvalue the hard way	955
CPMS3	Characteristic polynomial of a matrix, size 3	961

Eigenvalues of a matrix, size 3	962
Eigenspaces of a matrix, size 3	966
Eigenvalue multiplicities, matrix of size 4	969
Eigenvalues, symmetric matrix of size 4	970
High multiplicity eigenvalues, matrix of size 5	973
Complex eigenvalues, matrix of size 6	974
Distinct eigenvalues, matrix of size 5	979
Έ	
Building desired eigenvalues	1008
Similar matrices of size 5	1030
Similar matrices of size 3	1031
Equal eigenvalues, not similar	1036
Diagonalization of Archetype B	1037
Diagonalizing a matrix of size 3	1042
A non-diagonalizable matrix of size 4	1047
Distinct eigenvalues, hence diagonalizable	1049
High power of a diagonalizable matrix	1051
Fibonacci sequence, closed form	1054
	Eigenspaces of a matrix, size 3

Section LT

	A linear transformation
NLT	Not a linear transformation
LTPM	Linear transformation, polynomials to matrices
LTPP	Linear transformation, polynomials to polynomials
LTM	Linear transformation from a matrix
MFLT	Matrix from a linear transformation
MOLT	Matrix of a linear transformation
LTDB1	Linear transformation defined on a basis
LTDB2	Linear transformation defined on a basis
LTDB3	Linear transformation defined on a basis
SPIAS	Sample pre-images, Archetype S
STLT	Sum of two linear transformations
SMLT	Scalar multiple of a linear transformation
CTLT	Composition of two linear transformations

NIAQ	Not injective, Archetype Q	1129
IAR	Injective, Archetype R	1131
IAV		1135
NKAO	Nontrivial kernel, Archetype O	1138
TKAP	Trivial kernel, Archetype P	1141
NIAQR	Not injective, Archetype Q, revisited	1145
NIAO	Not injective, Archetype O	1147

IAP	Injective, Archetype P
NIDAU	Not injective by dimension, Archetype U 1151
Section SL	Г
NSAQ	Not surjective, Archetype Q
SAR	Surjective, Archetype R
SAV	Surjective, Archetype V
RAO	Range, Archetype O1178
FRAN	Full range, Archetype N 1182
NSAQR	Not surjective, Archetype Q, revisited
NSAO	Not surjective, Archetype O
SAN	Surjective, Archetype N
BRLT	A basis for the range of a linear transformation
NSDAT	Not surjective by dimension, Archetype T
Section IV	LT
AIVLT	An invertible linear transformation
ANILT	A non-invertible linear transformation
CIVLT	Computing the Inverse of a Linear Transformations
IVSAV	Isomorphic vector spaces, Archetype V 1230
Section VF	
VRC4	Vector representation in \mathbb{C}^4

VRP2	Vector representations in P_2	1267				
TIVS	Two isomorphic vector spaces	1273				
CVSR	Crazy vector space revealed	1273				
ASC	A subspace characterized	1273				
MIVS	Multiple isomorphic vector spaces	1274				
CP2	Coordinatizing in P_2	1277				
CM32	Coordinatization in M_{32}	1279				
Section MI	R					
OLTTR	One linear transformation, three representations	1286				
ALTMM	A linear transformation as matrix multiplication					
MPMR	Matrix product of matrix representations	1304				
KVMR	Kernel via matrix representation	1313				
RVMR	Range via matrix representation	1319				
ILTVR	Inverse of a linear transformation via a representation	1324				
Section CE	3					

ELTBM	Eigenvectors of linear transformation between matrices	1356
ELTBP	Eigenvectors of linear transformation between polynomials	1357
CBP	Change of basis with polynomials	1361
CBCV	Change of basis with column vectors	1367
MRCM	Matrix representations and change-of-basis matrices	1372
MRBE	Matrix representation with basis of eigenvectors	1378

ELTT	Eigenvectors of a linear transformation, twice	1386
CELT	Complex eigenvectors of a linear transformation	1397
Section OI)	
ANM	A normal matrix	1428
Section NI		
NM64	Nilpotent matrix, size 6, index 4	1437
NM62	Nilpotent matrix, size 6, index 2	1439
JB4	Jordan block, size 4	1441
NJB5	Nilpotent Jordan block, size 5	1442
NM83	Nilpotent matrix, size 8, index 3	1444
KPNLT	Kernels of powers of a nilpotent linear transformation	1454
CFNLT	Canonical form for a nilpotent linear transformation	1468
Section IS		
TIS	Two invariant subspaces	1476
EIS	Eigenspaces as invariant subspaces	1480
ISJB	Invariant subspaces and Jordan blocks	1482
GE4	Generalized eigenspaces, dimension 4 domain	1488
GE6	Generalized eigenspaces, dimension 6 domain	1490
LTRGE	Linear transformation restriction on generalized eigenspace	1495
ISMR4	Invariant subspaces, matrix representation, dimension 4 domain	1501

ISMR6 GENR6	Invariant subspaces, matrix representation, dimension 6 domain1Generalized eigenspaces and nilpotent restrictions, dimension 6 domain1						
Section JC							
JCF10	Jordan canonical form, size 10	1535					
Section CN	10						
ACN	Arithmetic of complex numbers	1595					
CSCN	Conjugate of some complex numbers	1599					
MSCN	Modulus of some complex numbers	1601					
Section SE	T						
SETM	Set membership	1602					
SSET	Subset	1603					
\mathbf{CS}	Cardinality and Size	1606					
SU	Set union	1607					
SI	Set intersection	1608					
\mathbf{SC}	Set complement	1609					
Section PT Section F							
IM11	Integers mod 11	1833					
VSIM5	Vector space over integers mod 5	1834					

SM2Z7 FF8	Symmetric matrices of size 2 over \mathbb{Z}_7	$1836 \\ 1837$
ГГО		1007
Section T		
Section HF)	
HP	Hadamard Product	1858
Section VN	M	
VM4	Vandermonde matrix of size 4	1870
Section PS	$^{\circ}M$	
Section RC)D	
ROD2	Rank one decomposition, size 2	1891
ROD4	Rank one decomposition, size 4	1893
Section TI)	
TD4	Triangular decomposition, size 4	1902
TDSSE	Triangular decomposition solves a system of equations	1905
TDEE6	Triangular decomposition, entry by entry, size 6	1911
Section SV	Ď	
Section SR		
Section PC)D	

Section CF		
PTFP	Polynomial through five points	1941
Section SA	S	
SS6W	Sharing a secret 6 ways	1952

Preface

This textbook is designed to teach the university mathematics student the basics of linear algebra and the techniques of formal mathematics. There are no prerequisites other than ordinary algebra, but it is probably best used by a student who has the "mathematical maturity" of a sophomore or junior. The text has two goals: to teach the fundamental concepts and techniques of matrix algebra and abstract vector spaces, and to teach the techniques associated with understanding the definitions and theorems forming a coherent area of mathematics. So there is an emphasis on worked examples of nontrivial size and on proving theorems carefully.

This book is copyrighted. This means that governments have granted the author a monopoly — the exclusive right to control the making of copies and derivative works for many years (too many years in some cases). It also gives others limited rights, generally referred to as "fair use," such as the right to quote sections in a review without seeking permission. However, the author licenses this book to anyone under the terms of the GNU

Free Documentation License (GFDL), which gives you more rights than most copyrights (see Appendix GFDL [1814]). Loosely speaking, you may make as many copies as you like at no cost, and you may distribute these unmodified copies if you please. You may modify the book for your own use. The catch is that if you make modifications and you distribute the modified version, or make use of portions in excess of fair use in another work, then you must also license the new work with the GFDL. So the book has lots of inherent freedom, and no one is allowed to distribute a derivative work that restricts these freedoms. (See the license itself in the appendix for the exact details of the additional rights you have been given.)

Notice that initially most people are struck by the notion that this book is **free** (the French would say *gratuit*, at no cost). And it is. However, it is more important that the book has **freedom** (the French would say *liberté*, liberty). It will never go "out of print" nor will there ever be trivial updates designed only to frustrate the used book market. Those considering teaching a course with this book can examine it thoroughly in advance. Adding new exercises or new sections has been purposely made very easy, and the hope is that others will contribute these modifications back for incorporation into the book, for the benefit of all.

Depending on how you received your copy, you may want to check for the latest version (and other news) at http://linear.ups.edu/.

Topics The first half of this text (through Chapter M [425]) is basically a course in matrix algebra, though the foundation of some more advanced ideas is also being formed in these early sections. Vectors are presented exclusively as column vectors (since we also have the typographic freedom to avoid writing a column vector inline as the transpose of a row vector),

and linear combinations are presented very early. Spans, null spaces, column spaces and row spaces are also presented early, simply as sets, saving most of their vector space properties for later, so they are familiar objects before being scrutinized carefully.

You cannot do *everything* early, so in particular matrix multiplication comes later than usual. However, with a definition built on linear combinations of column vectors, it should seem more natural than the more frequent definition using dot products of rows with columns. And this delay emphasizes that linear algebra is built upon vector addition and scalar multiplication. Of course, matrix inverses must wait for matrix multiplication, but this does not prevent nonsingular matrices from occurring sooner. Vector space properties are hinted at when vector and matrix operations are first defined, but the notion of a vector space is saved for a more axiomatic treatment later (Chapter VS [658]). Once bases and dimension have been explored in the context of vector spaces, linear transformations and their matrix representation follow. The goal of the book is to go as far as Jordan canonical form in the Core (Part C [2]), with less central topics collected in the Topics (Part T [1829]). A third part contains contributed applications (Part A [1940]), with notation and theorems integrated with the earlier two parts.

Linear algebra is an ideal subject for the novice mathematics student to learn how to develop a topic precisely, with all the rigor mathematics requires. Unfortunately, much of this rigor seems to have escaped the standard calculus curriculum, so for many university students this is their first exposure to careful definitions and theorems, and the expectation that they fully understand them, to say nothing of the expectation that they become proficient in formulating their own proofs. We have tried to make this text as helpful as possible with this transition. Every definition is stated carefully, set apart from the text. Likewise, every theorem is carefully stated, and almost every one has a complete proof. Theorems usually have just one conclusion, so they can be referenced precisely later. Definitions and theorems are cataloged in order of their appearance in the front of the book (Definitions [ix], Theorems [x]), and alphabetical order in the index at the back. Along the way, there are discussions of some more important ideas relating to formulating proofs (Proof Techniques [??]), which is part advice and part logic.

Origin and History This book is the result of the confluence of several related events and trends.

- At the University of Puget Sound we teach a one-semester, post-calculus linear algebra course to students majoring in mathematics, computer science, physics, chemistry and economics. Between January 1986 and June 2002, I taught this course seventeen times. For the Spring 2003 semester, I elected to convert my course notes to an electronic form so that it would be easier to incorporate the inevitable and nearly-constant revisions. Central to my new notes was a collection of stock examples that would be used repeatedly to illustrate new concepts. (These would become the Archetypes, Appendix A [1637].) It was only a short leap to then decide to distribute copies of these notes and examples to the students in the two sections of this course. As the semester wore on, the notes began to look less like notes and more like a textbook.
- I used the notes again in the Fall 2003 semester for a single section of the course. Simultaneously, the textbook I was using came out in a fifth edition. A new chapter was added toward the start of the book, and a few additional exercises were added in other

chapters. This demanded the annoyance of reworking my notes and list of suggested exercises to conform with the changed numbering of the chapters and exercises. I had an almost identical experience with the third course I was teaching that semester. I also learned that in the next academic year I would be teaching a course where my textbook of choice had gone out of print. I felt there had to be a better alternative to having the organization of my courses buffeted by the economics of traditional textbook publishing.

- I had used T_EX and the Internet for many years, so there was little to stand in the way of typesetting, distributing and "marketing" a free book. With recreational and professional interests in software development, I had long been fascinated by the open-source software movement, as exemplified by the success of GNU and Linux, though public-domain T_EX might also deserve mention. Obviously, this book is an attempt to carry over that model of creative endeavor to textbook publishing.
- As a sabbatical project during the Spring 2004 semester, I embarked on the current project of creating a freely-distributable linear algebra textbook. (Notice the implied financial support of the University of Puget Sound to this project.) Most of the material was written from scratch since changes in notation and approach made much of my notes of little use. By August 2004 I had written half the material necessary for our Math 232 course. The remaining half was written during the Fall 2004 semester as I taught another two sections of Math 232.
- While in early 2005 the book was complete enough to build a course around and Version

1.0 was released. Work has continued since, filling out the narrative, exercises and supplements.

However, much of my motivation for writing this book is captured by the sentiments expressed by H.M. Cundy and A.P. Rollet in their Preface to the First Edition of *Mathematical Models* (1952), especially the final sentence,

This book was born in the classroom, and arose from the spontaneous interest of a Mathematical Sixth in the construction of simple models. A desire to show that even in mathematics one could have fun led to an exhibition of the results and attracted considerable attention throughout the school. Since then the Sherborne collection has grown, ideas have come from many sources, and widespread interest has been shown. It seems therefore desirable to give permanent form to the lessons of experience so that others can benefit by them and be encouraged to undertake similar work.

How To Use This Book Chapters, Theorems, etc. are not numbered in this book, but are instead referenced by acronyms. This means that Theorem XYZ will always be Theorem XYZ, no matter if new sections are added, or if an individual decides to remove certain other sections. Within sections, the subsections are acronyms that begin with the acronym of the section. So Subsection XYZ.AB is the subsection AB in Section XYZ. Acronyms are unique within their type, so for example there is just one Definition B [777], but there is also a Section B [777]. At first, all the letters flying around may be confusing, but with time, you will begin to recognize the more important ones on sight. Furthermore, there are

lists of theorems, examples, etc. in the front of the book, and an index that contains every acronym. If you are reading this in an electronic version (PDF or XML), you will see that all of the cross-references are hyperlinks, allowing you to click to a definition or example, and then use the back button to return. In printed versions, you must rely on the page numbers. However, note that page numbers are not permanent! Different editions, different margins, or different sized paper will affect what content is on each page. And in time, the addition of new material will affect the page numbering.

Chapter divisions are not critical to the organization of the book, as Sections are the main organizational unit. Sections are designed to be the subject of a single lecture or classroom session, though there is frequently more material than can be discussed and illustrated in a fifty-minute session. Consequently, the instructor will need to be selective about which topics to illustrate with other examples and which topics to leave to the student's reading. Many of the examples are meant to be large, such as using five or six variables in a system of equations, so the instructor may just want to "walk" a class through these examples. The book has been written with the idea that some may work through it independently, so the hope is that students can learn some of the more mechanical ideas on their own.

The highest level division of the book is the three Parts: Core, Topics, Applications (Part C [2], Part T [1829], Part A [1940]). The Core is meant to carefully describe the basic ideas required of a first exposure to linear algebra. In the final sections of the Core, one should ask the question: which previous Sections could be removed without destroying the logical development of the subject? Hopefully, the answer is "none." The goal of the book is to finish the Core with a very general representation of a linear transformation (Jordan canonical form, Section JCF [1517]). Of course, there will not be universal agreement on

what should, or should not, constitute the Core, but the main idea is to limit it to about forty sections. Topics (Part T [1829]) is meant to contain those subjects that are important in linear algebra, and which would make profitable detours from the Core for those interested in pursuing them. Applications (Part A [1940]) should illustrate the power and widespread applicability of linear algebra to as many fields as possible. The Archetypes (Appendix A [1637]) cover many of the computational aspects of systems of linear equations, matrices and linear transformations. The student should consult them often, and this is encouraged by exercises that simply suggest the right properties to examine at the right time. But what is more important, this a repository that contains enough variety to provide abundant examples of key theorems, while also providing counterexamples to hypotheses or converses of theorems. The summary table at the start of this appendix should be especially useful.

I require my students to read each Section *prior* to the day's discussion on that section. For some students this is a novel idea, but at the end of the semester a few always report on the benefits, both for this course and other courses where they have adopted the habit. To make good on this requirement, each section contains three Reading Questions. These sometimes only require parroting back a key definition or theorem, or they require performing a small example of a key computation, or they ask for musings on key ideas or new relationships between old ideas. Answers are emailed to me the evening before the lecture. Given the flavor and purpose of these questions, including solutions seems foolish.

Every chapter of Part C [2] ends with "Annotated Acronyms", a short list of critical theorems or definitions from that chapter. There are a variety of reasons for any one of these to have been chosen, and reading the short paragraphs after some of these might provide insight into the possibilities. An end-of-chapter review might usefully incorporate a close

reading of these lists.

Formulating interesting and effective exercises is as difficult, or more so, than building a narrative. But it is the place where a student really learns the material. As such, for the student's benefit, complete solutions should be given. As the list of exercises expands, the amount with solutions should similarly expand. Exercises and their solutions are referenced with a section name, followed by a dot, then a letter (C,M, or T) and a number. The letter 'C' indicates a problem that is mostly computational in nature, while the letter 'T' indicates a problem that is more theoretical in nature. A problem with a letter 'M' is somewhere in between (middle, mid-level, median, middling), probably a mix of computation and applications of theorems. So Solution MO.T13 [459] is a solution to an exercise in Section MO [426] that is theoretical in nature. The number '13' has no intrinsic meaning.

More on Freedom This book is freely-distributable under the terms of the GFDL, along with the underlying T_EX code from which the book is built. This arrangement provides many benefits unavailable with traditional texts.

• No cost, or low cost, to students. With no physical vessel (i.e. paper, binding), no transportation costs (Internet bandwidth being a negligible cost) and no marketing costs (evaluation and desk copies are free to all), anyone with an Internet connection can obtain it, and a teacher could make available paper copies in sufficient quantities for a class. The cost to print a copy is not insignificant, but is just a fraction of the cost of a traditional textbook when printing is handled by a print-on-demand service over the Internet. Students will not feel the need to sell back their book (nor should there

be much of a market for used copies), and in future years can even pick up a newer edition freely.

• Electronic versions of the book contain extensive hyperlinks. Specifically, most logical steps in proofs and examples include links back to the previous definitions or theorems that support that step. With whatever viewer you might be using (web browser, PDF reader) the "back" button can then return you to the middle of the proof you were studying. So even if you are reading a physical copy of this book, you can benefit from also working with an electronic version.

A traditional book, which the publisher is unwilling to distribute in an easily-copied electronic form, cannot offer this very intuitive and flexible approach to learning mathematics.

- The book will not go out of print. No matter what, a teacher can maintain their own copy and use the book for as many years as they desire. Further, the naming schemes for chapters, sections, theorems, etc. is designed so that the addition of new material will not break any course syllabi or assignment list.
- With many eyes reading the book and with frequent postings of updates, the reliability should become very high. Please report any errors you find that persist into the latest version.
- For those with a working installation of the popular typesetting program T_EX , the book has been designed so that it can be customized. Page layouts, presence of exercises,

solutions, sections or chapters can all be easily controlled. Furthermore, many variants of mathematical notation are achieved via T_EX macros. So by changing a single macro, one's favorite notation can be reflected throughout the text. For example, every transpose of a matrix is coded in the source as \transpose{A} , which when printed will yield A^t . However by changing the definition of $\transpose{}$, any desired alternative notation (superscript t, superscript T, superscript prime) will then appear throughout the text instead.

- The book has also been designed to make it easy for others to contribute material. Would you like to see a section on symmetric bilinear forms? Consider writing one and contributing it to one of the Topics chapters. Should there be more exercises about the null space of a matrix? Send me some. Historical Notes? Contact me, and we will see about adding those in also.
- You have no legal obligation to pay for this book. It has been licensed with no expectation that you pay for it. You do not even have a moral obligation to pay for the book. Thomas Jefferson (1743 1826), the author of the United States Declaration of Independence, wrote,

If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density in any point, and like the air in which we breathe, move, and have our physical being, incapable of confinement or exclusive appropriation.

> Letter to Isaac McPherson August 13, 1813

However, if you feel a royalty is due the author, or if you would like to encourage the author, or if you wish to show others that this approach to textbook publishing can also bring financial compensation, then donations are gratefully received. Moreover, non-financial forms of help can often be even more valuable. A simple note of encouragement, submitting a report of an error, or contributing some exercises or perhaps an entire section for the Topics or Applications are all important ways you can acknowledge the freedoms accorded to this work by the copyright holder and other contributors.

Conclusion Foremost, I hope that students find their time spent with this book profitable. I hope that instructors find it flexible enough to fit the needs of their course. And I hope

that everyone will send me their comments and suggestions, and also consider the myriad ways they can help (as listed on the book's website at http://linear.ups.edu).

Robert A. Beezer Tacoma, Washington July 2008

Acknowledgements

Many people have helped to make this book, and its freedoms, possible.

First, the time to create, edit and distribute the book has been provided implicitly and explicitly by the University of Puget Sound. A subbatical leave Spring 2004 and a course release in Spring 2007 are two obvious examples of explicit support. The latter was provided by support from the Lind-VanEnkevort Fund. The university has also provided clerical support, computer hardware, network servers and bandwidth. Thanks to Dean Kris Bartanen and the chair of the Mathematics and Computer Science Department, Professor Martin Jackson, for their support, encouragement and flexibility.

My colleagues in the Mathematics and Computer Science Department have graciously taught our introductory linear algebra course using preliminary versions and have provided valuable suggestions that have improved the book immeasurably. Thanks to Professor Martin Jackson (v0.30), Professor David Scott (v0.70) and Professor Bryan Smith (v0.70, 0.80, v1.00).

University of Puget Sound librarians Lori Ricigliano, Elizabeth Knight and Jeanne Kimura provided valuable advice on production, and interesting conversations about copyrights.

Many aspects of the book have been influenced by insightful questions and creative suggestions from the students who have labored through the book in our courses. For example, the flashcards with theorems and definitions are a direct result of a student suggestion. I will single out a handful of students have been especially adept at finding and reporting mathematically significant typographical errors: Jake Linenthal, Christie Su, Kim Le, Sarah McQuate, Andy Zimmer, Travis Osborne, Andrew Tapay, Mark Shoemaker, Tasha Underhill, Tim Zitzer, Elizabeth Million, and Steve Canfield.

I have tried to be as original as possible in the organization and presentation of this beautiful subject. However, I have been influenced by many years of teaching from another excellent textbook, *Introduction to Linear Algebra* by L.W. Johnson, R.D. Reiss and J.T. Arnold. When I have needed inspiration for the correct approach to particularly important proofs, I have learned to eventually consult two other textbooks. Sheldon Axler's *Linear Algebra Done Right* is a highly original exposition, while Ben Noble's *Applied Linear Algebra* frequently strikes just the right note between rigor and intuition. Noble's excellent book is highly recommended, even though its publication dates to 1969.

Conversion to various electronic formats have greatly depended on assistance from: Eitan Gurari, author of the powerful LATEX translator, tex4ht; Davide Cervone, author of jsMath; and Carl Witty, who advised and tested the Sony Reader format. Thanks to these individuals for their critical assistance.

General support and encouragement of free and affordable textbooks, in addition to specific promotion of this text, was provided by Nicole Allen, Textbook Advocate at Student

Public Interest Research Groups. Nicole was an early consumer of this material, back when it looked more like lecture notes than a textbook.

Finally, in every possible case, the production and distribution of this book has been accomplished with open-source software. The range of individuals and projects is far too great to pretend to list them all. The book's web site will someday maintain pointers to as many of these projects as possible. Part C Core

Chapter SLE Systems of Linear Equations

We will motivate our study of linear algebra by studying solutions to systems of linear equations. While the focus of this chapter is on the practical matter of how to find, and describe, these solutions, we will also be setting ourselves up for more theoretical ideas that will appear later.

Section WILA What is Linear Algebra?

Subsection LA "Linear" + "Algebra"

The subject of linear algebra can be partially explained by the meaning of the two terms comprising the title. "Linear" is a term you will appreciate better at the end of this course, and indeed, attaining this appreciation could be taken as one of the primary goals of this course. However for now, you can understand it to mean anything that is "straight" or "flat." For example in the xy-plane you might be accustomed to describing straight lines (is there any other kind?) as the set of solutions to an equation of the form y = mx + b, where the slope m and the y-intercept b are constants that together describe the line. In multivariate calculus, you may have discussed planes. Living in three dimensions, with coordinates described by triples (x, y, z), they can be described as the set of solutions to equations of the form ax + by + cz = d, where a, b, c, d are constants that together determine the plane. While we might describe planes as "flat," lines in three dimensions might be described as "straight." From a multivariate calculus course you will recall that lines are

sets of points described by equations such as x = 3t - 4, y = -7t + 2, z = 9t, where t is a parameter that can take on any value.

Another view of this notion of "flatness" is to recognize that the sets of points just described are solutions to equations of a relatively simple form. These equations involve addition and multiplication only. We will have a need for subtraction, and occasionally we will divide, but mostly you can describe "linear" equations as involving only addition and multiplication. Here are some examples of typical equations we will see in the next few sections:

$$2x + 3y - 4z = 13 \qquad 4x_1 + 5x_2 - x_3 + x_4 + x_5 = 0 \qquad 9a - 2b + 7c + 2d = -7$$

What we will not see are equations like:

$$xy + 5yz = 13$$
 $x_1 + x_2^3/x_4 - x_3x_4x_5^2 = 0$ $\tan(ab) + \log(c - d) = -7$

The exception will be that we will on occasion need to take a square root.

You have probably heard the word "algebra" frequently in your mathematical preparation for this course. Most likely, you have spent a good ten to fifteen years learning the algebra of the real numbers, along with some introduction to the very similar algebra of complex numbers (see Section CNO [1594]). However, there are many new algebras to learn and use, and likely linear algebra will be your second algebra. Like learning a second language, the necessary adjustments can be challenging at times, but the rewards are many. And it will make learning your third and fourth algebras even easier. Perhaps you have heard of "groups" and "rings" (or maybe you have studied them already), which are excellent examples of other algebras with very interesting properties and applications. In any event, prepare yourself to learn a new algebra and realize that some of the old rules you used for the real numbers may no longer apply to this *new* algebra you will be learning!

The brief discussion above about lines and planes suggests that linear algebra has an inherently geometric nature, and this is true. Examples in two and three dimensions can be used to provide valuable insight into important concepts of this course. However, much of the power of linear algebra will be the ability to work with "flat" or "straight" objects in higher dimensions, without concerning ourselves with visualizing the situation. While much of our intuition will come from examples in two and three dimensions, we will maintain an *algebraic* approach to the subject, with the geometry being secondary. Others may wish to switch this emphasis around, and that can lead to a very fruitful and beneficial course, but here and now we are laying our bias bare.

Subsection AA An Application

We conclude this section with a rather involved example that will highlight some of the power and techniques of linear algebra. Work through all of the details with pencil and paper, until you believe all the assertions made. However, in this introductory example, do not concern yourself with how some of the results are obtained or how you might be expected to solve a similar problem. We will come back to this example later and expose some of the techniques used and properties exploited. For now, use your background in mathematics to convince yourself that everything said here really is correct.

Example TMP Trail Mix Packaging

Suppose you are the production manager at a food-packaging plant and one of your product lines is trail mix, a healthy snack popular with hikers and backpackers, containing raisins, peanuts and hard-shelled chocolate pieces. By adjusting the mix of these three ingredients, you are able to sell three varieties of this item. The fancy version is sold in half-kilogram packages at outdoor supply stores and has more chocolate and fewer raisins, thus commanding a higher price. The standard version is sold in one kilogram packages in grocery stores and gas station mini-markets. Since the standard version has roughly equal amounts of each ingredient, it is not as expensive as the fancy version. Finally, a bulk version is sold in bins at grocery stores for consumers to load into plastic bags in amounts of their choosing. To appeal to the shoppers that like bulk items for their economy and healthfulness, this mix has many more raisins (at the expense of chocolate) and therefore sells for less.

Your production facilities have limited storage space and early each morning you are able to receive and store 380 kilograms of raisins, 500 kilograms of peanuts and 620 kilograms of chocolate pieces. As production manager, one of your most important duties is to decide how much of each version of trail mix to make every day. Clearly, you can have up to 1500 kilograms of raw ingredients available each day, so to be the most productive you will likely produce 1500 kilograms of trail mix each day. Also, you would prefer not to have any ingredients leftover each day, so that your final product is as fresh as possible and so that you can receive the maximum delivery the next morning. But how should these ingredients be allocated to the mixing of the bulk, standard and fancy versions?

First, we need a little more information about the mixes. Workers mix the ingredients in 15 kilogram batches, and each row of the table below gives a recipe for a 15 kilogram batch. There is some additional information on the costs of the ingredients and the price the manufacturer can charge for the different versions of the trail mix.

	Raisins	Peanuts	Chocolate	Cost	Sale Price
	(kg/batch)	(kg/batch)	(kg/batch)	(\$/kg)	(\$/kg)
Bulk	7	6	2	3.69	4.99
Standard	6	4	5	3.86	5.50
Fancy	2	5	8	4.45	6.50
Storage (kg)	380	500	620		
Cost $(\$/kg)$	2.55	4.65	4.80		

As production manager, it is important to realize that you only have three decisions to make — the amount of bulk mix to make, the amount of standard mix to make and the amount of fancy mix to make. Everything else is beyond your control or is handled by another department within the company. Principally, you are also limited by the amount of raw ingredients you can store each day. Let us denote the amount of each mix to produce each day, measured in kilograms, by the variable quantities b, s and f. Your production schedule can be described as values of b, s and f that do several things. First, we cannot make

negative quantities of each mix, so

$$b \ge 0$$
 $s \ge 0$ $f \ge 0$

Second, if we want to consume all of our ingredients each day, the storage capacities lead to three (linear) equations, one for each ingredient,

$$\frac{7}{15}b + \frac{6}{15}s + \frac{2}{15}f = 380$$
 (raisins)
$$\frac{6}{15}b + \frac{4}{15}s + \frac{5}{15}f = 500$$
 (peanuts)
$$\frac{2}{15}b + \frac{5}{15}s + \frac{8}{15}f = 620$$
 (chocolate)

It happens that this system of three equations has just one solution. In other words, as production manager, your job is easy, since there is but one way to use up all of your raw ingredients making trail mix. This single solution is

$$b = 300 \text{ kg}$$
 $s = 300 \text{ kg}$ $f = 900 \text{ kg}$.

We do not yet have the tools to explain why this solution is the only one, but it should be simple for you to verify that this is indeed a solution. (Go ahead, we will wait.) Determining solutions such as this, and establishing that they are unique, will be the main motivation for our initial study of linear algebra. So we have solved the problem of making sure that we make the best use of our limited storage space, and each day use up all of the raw ingredients that are shipped to us. Additionally, as production manager, you must report weekly to the CEO of the company, and you know he will be more interested in the profit derived from your decisions than in the actual production levels. So you compute,

300(4.99 - 3.69) + 300(5.50 - 3.86) + 900(6.50 - 4.45) = 2727.00

for a daily profit of \$2,727 from this production schedule. The computation of the daily profit is also beyond our control, though it is definitely of interest, and it too looks like a "linear" computation.

As often happens, things do not stay the same for long, and now the marketing department has suggested that your company's trail mix products standardize on every mix being one-third peanuts. Adjusting the peanut portion of each recipe by also adjusting the chocolate portion leads to revised recipes, and slightly different costs for the bulk and standard mixes, as given in the following table.

	Raisins	Peanuts	Chocolate	Cost	Sale Price
	(kg/batch)	(kg/batch)	(kg/batch)	(\$/kg)	(\$/kg)
Bulk	7	5	3	3.70	4.99
Standard	6	5	4	3.85	5.50
Fancy	2	5	8	4.45	6.50
Storage (kg)	380	500	620		
Cost (\$/kg)	2.55	4.65	4.80		

In a similar fashion as before, we desire values of b, s and f so that

$$b \ge 0$$
 $s \ge 0$ $f \ge 0$

and

$$\frac{7}{15}b + \frac{6}{15}s + \frac{2}{15}f = 380$$
 (raisins)
$$\frac{5}{15}b + \frac{5}{15}s + \frac{5}{15}f = 500$$
 (peanuts)
$$\frac{3}{15}b + \frac{4}{15}s + \frac{8}{15}f = 620$$
 (chocolate)

It now happens that this system of equations has *infinitely* many solutions, as we will now demonstrate. Let f remain a variable quantity. Then if we make f kilograms of the fancy mix, we will make 4f - 3300 kilograms of the bulk mix and -5f + 4800 kilograms of the standard mix. Let us now verify that, for any choice of f, the values of b = 4f - 3300 and s = -5f + 4800 will yield a production schedule that exhausts all of the day's supply of raw ingredients (right now, do not be concerned about how you might derive expressions like these for b and s). Grab your pencil and paper and play along.

$$\frac{7}{15}(4f - 3300) + \frac{6}{15}(-5f + 4800) + \frac{2}{15}f = 0f + \frac{5700}{15} = 380$$
$$\frac{5}{15}(4f - 3300) + \frac{5}{15}(-5f + 4800) + \frac{5}{15}f = 0f + \frac{7500}{15} = 500$$

$$\frac{3}{15}(4f - 3300) + \frac{4}{15}(-5f + 4800) + \frac{8}{15}f = 0f + \frac{9300}{15} = 620$$

Convince yourself that these expressions for b and s allow us to vary f and obtain an infinite number of possibilities for solutions to the three equations that describe our storage capacities. As a practical matter, there really are not an infinite number of solutions, since we are unlikely to want to end the day with a fractional number of bags of fancy mix, so our allowable values of f should probably be integers. More importantly, we need to remember that we cannot make negative amounts of each mix! Where does this lead us? Positive quantities of the bulk mix requires that

 $b \ge 0 \quad \Rightarrow \quad 4f - 3300 \ge 0 \quad \Rightarrow \quad f \ge 825$

Similarly for the standard mix,

$$s \ge 0 \quad \Rightarrow \quad -5f + 4800 \ge 0 \quad \Rightarrow \quad f \le 960$$

So, as production manager, you really have to choose a value of f from the finite set

 $\{825, 826, \ldots, 960\}$

leaving you with 136 choices, each of which will exhaust the day's supply of raw ingredients. Pause now and think about which *you* would choose.

Recalling your weekly meeting with the CEO suggests that you might want to choose a production schedule that yields the biggest possible profit for the company. So you compute an expression for the profit based on your as yet undetermined decision for the value of f,

(4f - 3300)(4.99 - 3.70) + (-5f + 4800)(5.50 - 3.85) + (f)(6.50 - 4.45) = -1.04f + 3663

Since f has a negative coefficient it would appear that mixing fancy mix is detrimental to your profit and should be avoided. So you will make the decision to set daily fancy mix production at f = 825. This has the effect of setting b = 4(825) - 3300 = 0 and we stop producing bulk mix entirely. So the remainder of your daily production is standard mix at the level of s = -5(825) + 4800 = 675 kilograms and the resulting daily profit is (-1.04)(825) + 3663 = 2805. It is a pleasant surprise that daily profit has risen to \$2,805, but this is not the most important part of the story. What is important here is that there are a large number of ways to produce trail mix that use all of the day's worth of raw ingredients and you were able to easily choose the one that netted the largest profit. Notice too how all of the above computations look "linear."

In the food industry, things do not stay the same for long, and now the sales department says that increased competition has led to the decision to stay competitive and charge just \$5.25 for a kilogram of the standard mix, rather than the previous \$5.50 per kilogram. This decision has no effect on the possibilities for the production schedule, but will affect the decision based on profit considerations. So you revisit just the profit computation, suitably adjusted for the new selling price of standard mix,

(4f - 3300)(4.99 - 3.70) + (-5f + 4800)(5.25 - 3.85) + (f)(6.50 - 4.45) = 0.21f + 2463

Now it would appear that fancy mix is beneficial to the company's profit since the value of f has a positive coefficient. So you take the decision to make as much fancy mix as possible, setting f = 960. This leads to s = -5(960) + 4800 = 0 and the increased competition has driven you out of the standard mix market all together. The remainder of production is therefore bulk mix at a daily level of b = 4(960) - 3300 = 540 kilograms and the resulting

daily profit is 0.21(960) + 2463 = 2664.60. A daily profit of \$2,664.60 is less than it used to be, but as production manager, you have made the best of a difficult situation and shown the sales department that the best course is to pull out of the highly competitive standard mix market completely.

This example is taken from a field of mathematics variously known by names such as operations research, systems science, or management science. More specifically, this is a perfect example of problems that are solved by the techniques of "linear programming."

There is a lot going on under the hood in this example. The heart of the matter is the solution to systems of linear equations, which is the topic of the next few sections, and a recurrent theme throughout this course. We will return to this example on several occasions to reveal some of the reasons for its behavior.

Subsection READ Reading Questions

- 1. Is the equation $x^2 + xy + \tan(y^3) = 0$ linear or not? Why or why not?
- 2. Find all solutions to the system of two linear equations 2x + 3y = -8, x y = 6.
- 3. Describe how the production manager might explain the importance of the procedures described in the trail mix application (Subsection WILA.AA [5]).

Subsection EXC Exercises

C10 In Example TMP [6] the first table lists the cost (per kilogram) to manufacture each of the three varieties of trail mix (bulk, standard, fancy). For example, it costs \$3.69 to make one kilogram of the bulk variety. Re-compute each of these three costs and notice that the computations are linear in character.

Contributed by Robert Beezer

M70 In Example TMP [6] two different prices were considered for marketing standard mix with the revised recipes (one-third peanuts in each recipe). Selling standard mix at \$5.50 resulted in selling the minimum amount of the fancy mix and no bulk mix. At \$5.25 it was best for profits to sell the maximum amount of fancy mix and then sell no standard mix. Determine a selling price for standard mix that allows for maximum profits while still selling some of each type of mix.

Contributed by Robert Beezer Solution [15]

Subsection SOL Solutions

M70 Contributed by Robert Beezer Statement [14]

If the price of standard mix is set at \$5.292, then the profit function has a zero coefficient on the variable quantity f. So, we can set f to be any integer quantity in {825, 826, ..., 960}. All but the extreme values (f = 825, f = 960) will result in production levels where some of every mix is manufactured. No matter what value of f is chosen, the resulting profit will be the same, at \$2,664.60.

Section SSLE Solving Systems of Linear Equations

We will motivate our study of linear algebra by considering the problem of solving several linear equations simultaneously. The word "solve" tends to get abused somewhat, as in "solve this problem." When talking about equations we understand a more precise meaning: find *all* of the values of some variable quantities that make an equation, or several equations, true.

Subsection SLE Systems of Linear Equations

Example STNE Solving two (nonlinear) equations Suppose we desire the simultaneous solutions of the two equations,

$$x^2 + y^2 = 1$$
$$-x + \sqrt{3}y = 0$$

You can easily check by substitution that $x = \frac{\sqrt{3}}{2}$, $y = \frac{1}{2}$ and $x = -\frac{\sqrt{3}}{2}$, $y = -\frac{1}{2}$ are both solutions. We need to also convince ourselves that these are the *only* solutions. To see this, plot each equation on the *xy*-plane, which means to plot (x, y) pairs that make an individual equation true. In this case we get a circle centered at the origin with radius 1 and a straight line through the origin with slope $\frac{1}{\sqrt{3}}$. The intersections of these two curves are our desired simultaneous solutions, and so we believe from our plot that the two solutions we know already are indeed the only ones. We like to write solutions as sets, so in this case we write the set of solutions as

$$S = \left\{ \left(\frac{\sqrt{3}}{2}, \frac{1}{2}\right), \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right) \right\}$$

 \boxtimes

In order to discuss systems of linear equations carefully, we need a precise definition. And before we do that, we will introduce our periodic discussions about "Proof Techniques." Linear algebra is an excellent setting for learning how to read, understand and formulate proofs. But this is a difficult step in your development as a mathematician, so we have included a series of short essays containing advice and explanations to help you along. These can be found back in Section PT [1611] of Appendix P [1593], and we will reference them as they become appropriate. Be sure to head back to the appendix to read this as they are introduced. With a definition next, now is the time for the first of our proof techniques. Head back to Section PT [1611] of Appendix P [1593] and study Technique D [1611]. We'll be right here when you get back. See you in a bit.

Definition SLE System of Linear Equations

A system of linear equations is a collection of m equations in the variable quantities $x_1, x_2, x_3, \ldots, x_n$ of the form,

 $a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$ $a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$ $a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$ \vdots $a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$

where the values of a_{ij} , b_i and x_j are from the set of complex numbers, \mathbb{C} .

Don't let the mention of the complex numbers, \mathbb{C} , rattle you. We will stick with real numbers exclusively for many more sections, and it will sometimes seem like we only work with integers! However, we want to leave the possibility of complex numbers open, and there will be occasions in subsequent sections where they are necessary. You can review the basic properties of complex numbers in Section CNO [1594], but these facts will not be critical until we reach Section O [395].

Now we make the notion of a solution to a linear system precise.

Definition SSLE Solution of a System of Linear Equations

A solution of a system of linear equations in n variables, $x_1, x_2, x_3, \ldots, x_n$ (such as the

system given in Definition SLE [18], is an ordered list of n complex numbers, $s_1, s_2, s_3, \ldots, s_n$ such that if we substitute s_1 for x_1, s_2 for x_2, s_3 for x_3, \ldots, s_n for x_n , then for every equation of the system the left side will equal the right side, i.e. each equation is true simultaneously.

More typically, we will write a solution in a form like $x_1 = 12$, $x_2 = -7$, $x_3 = 2$ to mean that $s_1 = 12$, $s_2 = -7$, $s_3 = 2$ in the notation of Definition SSLE [18]. To discuss all of the possible solutions to a system of linear equations, we now define the set of all solutions. (So Section SET [1602] is now applicable, and you may want to go and familiarize yourself with what is there.)

Definition SSSLE

Solution Set of a System of Linear Equations

The solution set of a linear system of equations is the set which contains every solution to the system, and nothing more. \triangle

Be aware that a solution set can be infinite, or there can be no solutions, in which case we write the solution set as the empty set, $\emptyset = \{\}$ (Definition ES [1603]). Here is an example to illustrate using the notation introduced in Definition SLE [18] and the notion of a solution (Definition SSLE [18]).

Example NSE Notation for a system of equations Given the system of linear equations,

$$x_1 + 2x_2 + \quad x_4 = 7$$

 $x_1 + x_2 + x_3 - x_4 = 3$ $3x_1 + x_2 + 5x_3 - 7x_4 = 1$

we have n = 4 variables and m = 3 equations. Also,

$a_{11} = 1$	$a_{12} = 2$	$a_{13} = 0$	$a_{14} = 1$	$b_1 = 7$
$a_{21} = 1$	$a_{22} = 1$	$a_{23} = 1$	$a_{24} = -1$	$b_2 = 3$
$a_{31} = 3$	$a_{32} = 1$	$a_{33} = 5$	$a_{34} = -7$	$b_3 = 1$

Additionally, convince yourself that $x_1 = -2$, $x_2 = 4$, $x_3 = 2$, $x_4 = 1$ is one solution (Definition SSLE [18]), but it is not the only one! For example, another solution is $x_1 = -12$, $x_2 = 11$, $x_3 = 1$, $x_4 = -3$, and there are more to be found. So the solution set contains at least two elements.

We will often shorten the term "system of linear equations" to "system of equations" leaving the linear aspect implied. After all, this is a book about *linear* algebra.

Subsection PSS Possibilities for Solution Sets

The next example illustrates the possibilities for the solution set of a system of linear equations. We will not be too formal here, and the necessary theorems to back up our claims will come in subsequent sections. So read for feeling and come back later to revisit this example.

Example TTS Three typical systems

Consider the system of two equations with two variables,

$$2x_1 + 3x_2 = 3 x_1 - x_2 = 4$$

If we plot the solutions to each of these equations separately on the x_1x_2 -plane, we get two lines, one with negative slope, the other with positive slope. They have exactly one point in common, $(x_1, x_2) = (3, -1)$, which is the solution $x_1 = 3$, $x_2 = -1$. From the geometry, we believe that this is the only solution to the system of equations, and so we say it is unique.

Now adjust the system with a different second equation,

$$2x_1 + 3x_2 = 3 4x_1 + 6x_2 = 6$$

A plot of the solutions to these equations individually results in two lines, one on top of the other! There are infinitely many pairs of points that make both equations true. We will learn shortly how to describe this infinite solution set precisely (see Example SAA [81], Theorem VFSLS [244]). Notice now how the second equation is just a multiple of the first.

One more minor adjustment provides a third system of linear equations,

$$2x_1 + 3x_2 = 3 4x_1 + 6x_2 = 10$$

A plot now reveals two lines with identical slopes, i.e. parallel lines. They have no points in common, and so the system has a solution set that is empty, $S = \emptyset$.

This example exhibits all of the typical behaviors of a system of equations. A subsequent theorem will tell us that every system of linear equations has a solution set that is empty, contains a single solution or contains infinitely many solutions (Theorem PSSLS [126]). Example STNE [16] yielded exactly two solutions, but this does not contradict the forthcoming theorem. The equations in Example STNE [16] are not linear because they do not match the form of Definition SLE [18], and so we cannot apply Theorem PSSLS [126] in this case.

Subsection ESEO Equivalent Systems and Equation Operations

With all this talk about finding solution sets for systems of linear equations, you might be ready to begin learning how to find these solution sets yourself. We begin with our first definition that takes a common word and gives it a very precise meaning in the context of systems of linear equations.

Definition ESYS Equivalent Systems

Two systems of linear equations are **equivalent** if their solution sets are equal. \triangle

Notice here that the two systems of equations could *look* very different (i.e. not be equal),

but still have equal solution sets, and we would then call the systems equivalent. Two linear equations in two variables might be plotted as two lines that intersect in a single point. A different system, with three equations in two variables might have a plot that is three lines, all intersecting at a common point, with this common point identical to the intersection point for the first system. By our definition, we could then say these two very different looking systems of equations are equivalent, since they have identical solution sets. It is really like a weaker form of equality, where we allow the systems to be different in some respects, but we use the term equivalent to highlight the situation when their solution sets are equal.

With this definition, we can begin to describe our strategy for solving linear systems. Given a system of linear equations that looks difficult to solve, we would like to have an *equivalent* system that is easy to solve. Since the systems will have equal solution sets, we can solve the "easy" system and get the solution set to the "difficult" system. Here come the tools for making this strategy viable.

Definition EO Equation Operations

Given a system of linear equations, the following three operations will transform the system into a different one, and each operation is known as an **equation operation**.

- 1. Swap the locations of two equations in the list of equations.
- 2. Multiply each term of an equation by a nonzero quantity.
- 3. Multiply each term of one equation by some quantity, and add these terms to a second equation, on both sides of the equality. Leave the first equation the same after this

operation, but replace the second equation by the new one.

 \triangle

These descriptions might seem a bit vague, but the proof or the examples that follow should make it clear what is meant by each. We will shortly prove a key theorem about equation operations and solutions to linear systems of equations. We are about to give a rather involved proof, so a discussion about just what a theorem really is would be timely. Head back and read Technique T [1613]. In the theorem we are about to prove, the conclusion is that two systems are equivalent. By Definition ESYS [22] this translates to requiring that solution sets be equal for the two systems. So we are being asked to show *that two sets are equal*. How do we do this? Well, there is a very standard technique, and we will use it repeatedly through the course. If you have not done so already, head to Section SET [1602] and familiarize yourself with sets, their operations, and especially the notion of set equality, Definition SE [1604] and the nearby discussion about its use.

Theorem EOPSS

Equation Operations Preserve Solution Sets

If we apply one of the three equation operations of Definition EO [23] to a system of linear equations (Definition SLE [18]), then the original system and the transformed system are equivalent. \Box

Proof We take each equation operation in turn and show that the solution sets of the two systems are equal, using the definition of set equality (Definition SE [1604]).

1. It will not be our habit in proofs to resort to saying statements are "obvious," but

in this case, it should be. There is nothing about the *order* in which we write linear equations that affects their solutions, so the solution set will be equal if the systems only differ by a rearrangement of the order of the equations.

2. Suppose $\alpha \neq 0$ is a number. Let's choose to multiply the terms of equation *i* by α to build the new system of equations,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$\alpha a_{i1}x_1 + \alpha a_{i2}x_2 + \alpha a_{i3}x_3 + \dots + \alpha a_{in}x_n = \alpha b_i$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$$

Let S denote the solutions to the system in the statement of the theorem, and let T denote the solutions to the transformed system.

(a) Show $S \subseteq T$. Suppose $(x_1, x_2, x_3, \ldots, x_n) = (\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in S$ is a solution to the original system. Ignoring the *i*-th equation for a moment, we know it makes all the other equations of the transformed system true. We also know that

$$a_{i1}\beta_1 + a_{i2}\beta_2 + a_{i3}\beta_3 + \dots + a_{in}\beta_n = b_i$$

which we can multiply by α to get

$$\alpha a_{i1}\beta_1 + \alpha a_{i2}\beta_2 + \alpha a_{i3}\beta_3 + \dots + \alpha a_{in}\beta_n = \alpha b_i$$

This says that the *i*-th equation of the transformed system is also true, so we have established that $(\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in T$, and therefore $S \subseteq T$.

(b) Now show $T \subseteq S$. Suppose $(x_1, x_2, x_3, \ldots, x_n) = (\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in T$ is a solution to the transformed system. Ignoring the *i*-th equation for a moment, we know it makes all the other equations of the original system true. We also know that

$$\alpha a_{i1}\beta_1 + \alpha a_{i2}\beta_2 + \alpha a_{i3}\beta_3 + \dots + \alpha a_{in}\beta_n = \alpha b_i$$

which we can multiply by $\frac{1}{\alpha}$, since $\alpha \neq 0$, to get

$$a_{i1}\beta_1 + a_{i2}\beta_2 + a_{i3}\beta_3 + \dots + a_{in}\beta_n = b_i$$

This says that the *i*-th equation of the original system is also true, so we have established that $(\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in S$, and therefore $T \subseteq S$. Locate the key point where we required that $\alpha \neq 0$, and consider what would happen if $\alpha = 0$.

3. Suppose α is a number. Let's choose to multiply the terms of equation *i* by α and add them to equation *j* in order to build the new system of equations,

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$(\alpha a_{i1} + a_{j1})x_1 + (\alpha a_{i2} + a_{j2})x_2 + \dots + (\alpha a_{in} + a_{jn})x_n = \alpha b_i + b_j$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

Let S denote the solutions to the system in the statement of the theorem, and let T denote the solutions to the transformed system.

(a) Show $S \subseteq T$. Suppose $(x_1, x_2, x_3, \ldots, x_n) = (\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in S$ is a solution to the original system. Ignoring the *j*-th equation for a moment, we know this solution makes all the other equations of the transformed system true. Using the fact that the solution makes the *i*-th and *j*-th equations of the original system true, we find

$$(\alpha a_{i1} + a_{j1})\beta_1 + (\alpha a_{i2} + a_{j2})\beta_2 + \dots + (\alpha a_{in} + a_{jn})\beta_n$$

= $(\alpha a_{i1}\beta_1 + \alpha a_{i2}\beta_2 + \dots + \alpha a_{in}\beta_n) + (a_{j1}\beta_1 + a_{j2}\beta_2 + \dots + a_{jn}\beta_n)$
= $\alpha (a_{i1}\beta_1 + a_{i2}\beta_2 + \dots + a_{in}\beta_n) + (a_{j1}\beta_1 + a_{j2}\beta_2 + \dots + a_{jn}\beta_n)$
= $\alpha b_i + b_j.$

This says that the *j*-th equation of the transformed system is also true, so we have established that $(\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in T$, and therefore $S \subseteq T$.

(b) Now show $T \subseteq S$. Suppose $(x_1, x_2, x_3, \ldots, x_n) = (\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in T$ is a solution to the transformed system. Ignoring the *j*-th equation for a moment, we know it makes all the other equations of the original system true. We then find

$$\begin{aligned} a_{j1}\beta_1 + a_{j2}\beta_2 + \dots + a_{jn}\beta_n \\ &= a_{j1}\beta_1 + a_{j2}\beta_2 + \dots + a_{jn}\beta_n + \alpha b_i - \alpha b_i \\ &= a_{j1}\beta_1 + a_{j2}\beta_2 + \dots + a_{jn}\beta_n + (\alpha a_{i1}\beta_1 + \alpha a_{i2}\beta_2 + \dots + \alpha a_{in}\beta_n) - \alpha b_i \\ &= a_{j1}\beta_1 + \alpha a_{i1}\beta_1 + a_{j2}\beta_2 + \alpha a_{i2}\beta_2 + \dots + a_{jn}\beta_n + \alpha a_{in}\beta_n - \alpha b_i \\ &= (\alpha a_{i1} + a_{j1})\beta_1 + (\alpha a_{i2} + a_{j2})\beta_2 + \dots + (\alpha a_{in} + a_{jn})\beta_n - \alpha b_i \\ &= \alpha b_i + b_j - \alpha b_i \\ &= b_j \end{aligned}$$

This says that the *j*-th equation of the original system is also true, so we have established that $(\beta_1, \beta_2, \beta_3, \ldots, \beta_n) \in S$, and therefore $T \subseteq S$.

Why didn't we need to require that $\alpha \neq 0$ for this row operation? In other words, how does the third statement of the theorem read when $\alpha = 0$? Does our proof require some extra care when $\alpha = 0$? Compare your answers with the similar situation for the second row operation. (See Exercise SSLE.T20 [42].)

Theorem EOPSS [24] is the necessary tool to complete our strategy for solving systems of equations. We will use equation operations to move from one system to another, all the while keeping the solution set the same. With the right sequence of operations, we will arrive at a simpler equation to solve. The next two examples illustrate this idea, while saving some of the details for later.

Example US Three equations, one solution

We solve the following system by a sequence of equation operations.

 $x_1 + 2x_2 + 2x_3 = 4$ $x_1 + 3x_2 + 3x_3 = 5$ $2x_1 + 6x_2 + 5x_3 = 6$

 $\alpha = -1$ times equation 1, add to equation 2:

 $x_1 + 2x_2 + 2x_3 = 4$ $0x_1 + 1x_2 + 1x_3 = 1$ $2x_1 + 6x_2 + 5x_3 = 6$

 $\alpha = -2$ times equation 1, add to equation 3:

 $x_1 + 2x_2 + 2x_3 = 4$ $0x_1 + 1x_2 + 1x_3 = 1$ $0x_1 + 2x_2 + 1x_3 = -2$ $\alpha = -2$ times equation 2, add to equation 3:

$$x_1 + 2x_2 + 2x_3 = 4$$

$$0x_1 + 1x_2 + 1x_3 = 1$$

$$0x_1 + 0x_2 - 1x_3 = -4$$

 $\alpha = -1$ times equation 3:

$$x_1 + 2x_2 + 2x_3 = 4$$

$$0x_1 + 1x_2 + 1x_3 = 1$$

$$0x_1 + 0x_2 + 1x_3 = 4$$

which can be written more clearly as

$$x_1 + 2x_2 + 2x_3 = 4$$

 $x_2 + x_3 = 1$
 $x_3 = 4$

This is now a very easy system of equations to solve. The third equation requires that $x_3 = 4$ to be true. Making this substitution into equation 2 we arrive at $x_2 = -3$, and finally, substituting these values of x_2 and x_3 into the first equation, we find that $x_1 = 2$. Note too that this is the only solution to this final system of equations, since we were forced to

choose these values to make the equations true. Since we performed equation operations on each system to obtain the next one in the list, all of the systems listed here are all equivalent to each other by Theorem EOPSS [24]. Thus $(x_1, x_2, x_3) = (2, -3, 4)$ is the unique solution to the *original* system of equations (and all of the other intermediate systems of equations listed as we transformed one into another).

Example IS

Three equations, infinitely many solutions

The following system of equations made an appearance earlier in this section (Example NSE [19]), where we listed *one* of its solutions. Now, we will try to find all of the solutions to this system. Don't concern yourself too much about why we choose this particular sequence of equation operations, just believe that the work we do is all correct.

$$x_1 + 2x_2 + 0x_3 + x_4 = 7$$

$$x_1 + x_2 + x_3 - x_4 = 3$$

$$3x_1 + x_2 + 5x_3 - 7x_4 = 1$$

 $\alpha = -1$ times equation 1, add to equation 2:

$$x_1 + 2x_2 + 0x_3 + x_4 = 7$$

$$0x_1 - x_2 + x_3 - 2x_4 = -4$$

$$3x_1 + x_2 + 5x_3 - 7x_4 = 1$$

 $\alpha = -3$ times equation 1, add to equation 3:

$$x_1 + 2x_2 + 0x_3 + x_4 = 7$$

$$0x_1 - x_2 + x_3 - 2x_4 = -4$$

$$0x_1 - 5x_2 + 5x_3 - 10x_4 = -20$$

 $\alpha = -5$ times equation 2, add to equation 3:

$$x_1 + 2x_2 + 0x_3 + x_4 = 7$$

$$0x_1 - x_2 + x_3 - 2x_4 = -4$$

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 0$$

 $\alpha = -1$ times equation 2:

$$x_1 + 2x_2 + 0x_3 + x_4 = 7$$

$$0x_1 + x_2 - x_3 + 2x_4 = 4$$

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 0$$

 $\alpha = -2$ times equation 2, add to equation 1:

$$x_1 + 0x_2 + 2x_3 - 3x_4 = -1$$

$$0x_1 + x_2 - x_3 + 2x_4 = 4$$

$$0x_1 + 0x_2 + 0x_3 + 0x_4 = 0$$

which can be written more clearly as

$$x_1 + 2x_3 - 3x_4 = -1$$
$$x_2 - x_3 + 2x_4 = 4$$
$$0 = 0$$

What does the equation 0 = 0 mean? We can choose *any* values for x_1, x_2, x_3, x_4 and this equation will be true, so we only need to consider further the first two equations, since the third is true no matter what. We can analyze the second equation without consideration of the variable x_1 . It would appear that there is considerable latitude in how we can choose x_2, x_3, x_4 and make this equation true. Let's choose x_3 and x_4 to be *anything* we please, say $x_3 = a$ and $x_4 = b$.

Now we can take these arbitrary values for x_3 and x_4 , substitute them in equation 1, to obtain

$$x_1 + 2a - 3b = -1$$
$$x_1 = -1 - 2a + 3b$$

Similarly, equation 2 becomes

$$x_2 - a + 2b = 4$$

$$x_2 = 4 + a - 2b$$

So our arbitrary choices of values for x_3 and x_4 (a and b) translate into specific values of x_1 and x_2 . The lone solution given in Example NSE [19] was obtained by choosing a = 2 and b = 1. Now we can easily and quickly find many more (infinitely more). Suppose we choose a = 5 and b = -2, then we compute

$$x_1 = -1 - 2(5) + 3(-2) = -17$$

$$x_2 = 4 + 5 - 2(-2) = 13$$

and you can verify that $(x_1, x_2, x_3, x_4) = (-17, 13, 5, -2)$ makes all three equations true. The entire solution set is written as

$$S = \{ (-1 - 2a + 3b, 4 + a - 2b, a, b) | a \in \mathbb{C}, b \in \mathbb{C} \}$$

It would be instructive to finish off your study of this example by taking the general form of the solutions given in this set and substituting them into each of the three equations and verify that they are true in each case (Exercise SSLE.M40 [41]). \boxtimes

In the next section we will describe how to use equation operations to systematically solve any system of linear equations. But first, read one of our more important pieces of advice about speaking and writing mathematics. See Technique L [1614].

Before attacking the exercises in this section, it will be helpful to read some advice on getting started on the construction of a proof. See Technique GS [1617].

Subsection READ Reading Questions

- 1. How many solutions does the system of equations 3x + 2y = 4, 6x + 4y = 8 have? Explain your answer.
- 2. How many solutions does the system of equations 3x + 2y = 4, 6x + 4y = -2 have? Explain your answer.
- 3. What do we mean when we say mathematics is a language?

Subsection EXC Exercises

C10 Find a solution to the system in Example IS [31] where $x_3 = 6$ and $x_4 = 2$. Find two other solutions to the system. Find a solution where $x_1 = -17$ and $x_2 = 14$. How many possible answers are there to each of these questions? Contributed by Robert Beezer

C20 Each archetype (Appendix A [1637]) that is a system of equations begins by listing some specific solutions. Verify the specific solutions listed in the following archetypes by evaluating the system of equations with the solutions listed.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670] Archetype E [1678] Archetype F [1686] Archetype G [1697] Archetype H [1705] Archetype I [1714] Archetype J [1723] Contributed by Robert Beezer C30 Find all solutions to the linear system:

$$\begin{aligned} x + y &= 5\\ 2x - y &= 3 \end{aligned}$$

Contributed by Chris Black Solution [43]

C31 Find all solutions to the linear system:

$$3x + 2y = 1$$
$$x - y = 2$$
$$4x + 2y = 2$$

Contributed by Chris Black

C32 Find all solutions to the linear system:

$$x + 2y = 8$$
$$x - y = 2$$
$$x + y = 4$$

Contributed by Chris Black

C33 Find all solutions to the linear system:

$$x + y - z = -1$$
$$x - y - z = -1$$
$$z = 2$$

Contributed by Chris Black

C34 Find all solutions to the linear system:

$$x + y - z = -5$$
$$x - y - z = -3$$
$$x + y - z = 0$$

Contributed by Chris Black

C50 A three-digit number has two properties. The tens-digit and the ones-digit add up to 5. If the number is written with the digits in the reverse order, and then subtracted from the original number, the result is 792. Use a system of equations to find all of the three-digit numbers with these properties.

Contributed by Robert Beezer Solution [43]

C51 Find all of the six-digit numbers in which the first digit is one less than the second, the third digit is half the second, the fourth digit is three times the third and the last two

digits form a number that equals the sum of the fourth and fifth. The sum of all the digits is 24. (From *The MENSA Puzzle Calendar* for January 9, 2006.) Contributed by Robert Beezer Solution [44]

C52 Driving along, Terry notices that the last four digits on his car's odometer are palindromic. A mile later, the last five digits are palindromic. After driving another mile, the middle four digits are palindromic. One more mile, and all six are palindromic. What was the odometer reading when Terry first looked at it? Form a linear system of equations that expresses the requirements of this puzzle. (*Car Talk* Puzzler, National Public Radio, Week of January 21, 2008) (A car odometer displays six digits and a sequence is a **palindrome** if it reads the same left-to-right as right-to-left.) Contributed by Robert Beezer Solution [45]

M10 Each sentence below has at least two meanings. Identify the source of the double meaning, and rewrite the sentence (at least twice) to clearly convey each meaning.

- 1. They are baking potatoes.
- 2. He bought many ripe pears and apricots.
- 3. She likes his sculpture.
- 4. I decided on the bus.

M11 Discuss the difference in meaning of each of the following three almost identical sentences, which all have the same grammatical structure. (These are due to Keith Devlin.)

1. She saw him in the park with a dog.

2. She saw him in the park with a fountain.

3. She saw him in the park with a telescope.

Contributed by Robert Beezer Solution [47]

M12 The following sentence, due to Noam Chomsky, has a correct grammatical structure, but is meaningless. Critique its faults. "Colorless green ideas sleep furiously." (Chomsky, Noam. Syntactic Structures, The Hague/Paris: Mouton, 1957. p. 15.) Contributed by Robert Beezer Solution [47]

M13 Read the following sentence and form a mental picture of the situation.

The baby cried and the mother picked it up.

What assumptions did you make about the situation? Contributed by Robert Beezer Solution [47]

M14 Discuss the difference in meaning of the following two almost identical sentences, which have nearly identical grammatical structure. (This antanaclasis is often attributed to the comedian Groucho Marx, but has earlier roots.)

- 1. Time flies like an arrow.
- 2. Fruit flies like a banana.

Contributed by Robert Beezer

M30 This problem appears in a middle-school mathematics textbook: Together Dan and Diane have \$20. Together Diane and Donna have \$15. How much do the three of them have in total? (*Transition Mathematics*, Second Edition, Scott Foresman Addison Wesley, 1998. Problem 5–1.19.)

Contributed by David Beezer Solution [47]

M40 Solutions to the system in Example IS [31] are given as

$$(x_1, x_2, x_3, x_4) = (-1 - 2a + 3b, 4 + a - 2b, a, b)$$

Evaluate the three equations of the original system with these expressions in a and b and verify that each equation is true, no matter what values are chosen for a and b. Contributed by Robert Beezer

M70 We have seen in this section that systems of linear equations have limited possibilities for solution sets, and we will shortly prove Theorem PSSLS [126] that describes these possibilities exactly. This exercise will show that if we relax the requirement that our equations be linear, then the possibilities expand greatly. Consider a system of two equations in the two variables x and y, where the departure from linearity involves simply squaring the variables.

$$x^2 - y^2 = 1$$
$$x^2 + y^2 = 4$$

After solving this system of *non-linear* equations, replace the second equation in turn by $x^2 + 2x + y^2 = 3$, $x^2 + y^2 = 1$, $x^2 - 4x + y^2 = -3$, $-x^2 + y^2 = 1$ and solve each resulting system of two equations in two variables. (This exercise includes suggestions from Don Kreher.)

Contributed by Robert Beezer Solution [48]

T10 Technique D [1611] asks you to formulate a definition of what it means for a whole number to be odd. What is your definition? (Don't say "the opposite of even.") Is 6 odd? Is 11 odd? Justify your answers by using your definition. Contributed by Robert Beezer Solution [48]

T20 Explain why the second equation operation in Definition EO [23] requires that the scalar be nonzero, while in the third equation operation this restriction on the scalar is not present.

Contributed by Robert Beezer Solution [48]

Subsection SOL Solutions

C30 Contributed by Chris Black Statement [37] Solving each equation for y, we have the equivalent system

$$y = 5 - x$$
$$y = 2x - 3.$$

Setting these expressions for y equal, we have the equation 5 - x = 2x - 3, which quickly leads to $x = \frac{8}{3}$. Substituting for x in the first equation, we have $y = 5 - x = 5 - \frac{8}{3} = \frac{7}{3}$. Thus, the solution is $x = \frac{8}{3}$, $y = \frac{7}{3}$.

C50 Contributed by Robert Beezer Statement [38] Let *a* be the hundreds digit, *b* the tens digit, and *c* the ones digit. Then the first condition says that b + c = 5. The original number is 100a + 10b + c, while the reversed number is 100c + 10b + a. So the second condition is

$$792 = (100a + 10b + c) - (100c + 10b + a) = 99a - 99c$$

So we arrive at the system of equations

$$b+c=5$$

$$99a - 99c = 792$$

Using equation operations, we arrive at the equivalent system

$$a - c = 8$$
$$b + c = 5$$

We can vary c and obtain infinitely many solutions. However, c must be a digit, restricting us to ten values (0 - 9). Furthermore, if c > 1, then the first equation forces a > 9, an impossibility. Setting c = 0, yields 850 as a solution, and setting c = 1 yields 941 as another solution.

C51 Contributed by Robert Beezer Statement [38] Let abcdef denote any such six-digit number and convert each requirement in the problem statement into an equation.

$$a = b - 1$$

$$c = \frac{1}{2}b$$

$$d = 3c$$

$$10e + f = d + e$$

$$24 = a + b + c + d + e + f$$

In a more standard form this becomes

a - b = -1

$$-b + 2c = 0$$
$$-3c + d = 0$$
$$-d + 9e + f = 0$$
$$a + b + c + d + e + f = 24$$

Using equation operations (or the techniques of the upcoming Section RREF [50]), this system can be converted to the equivalent system

$$a + \frac{16}{75}f = 5$$

$$b + \frac{16}{75}f = 6$$

$$c + \frac{8}{75}f = 3$$

$$d + \frac{8}{25}f = 9$$

$$e + \frac{11}{75}f = 1$$

Clearly, choosing f = 0 will yield the solution abcde = 563910. Furthermore, to have the variables result in single-digit numbers, none of the other choices for f (1, 2, ..., 9) will yield a solution.

C52 Contributed by Robert Beezer Statement [39] 198888 is one solution, and David Braithwaite found 199999 as another.

M10 Contributed by Robert Beezer Statement [39]

- Does "baking" describe the potato or what is happening to the potato? Those are potatoes that are used for baking. The potatoes are being baked.
- 2. Are the apricots ripe, or just the pears? Parentheses could indicate just what the adjective "ripe" is meant to modify. Were there many apricots as well, or just many pears? He bought many pears and many ripe apricots. He bought apricots and many ripe pears.
- Is "sculpture" a single physical object, or the sculptor's style expressed over many pieces and many years? She likes his sculpture of the girl. She likes his sculptural style.
- 4. Was a decision made while in the bus, or was the outcome of a decision to choose the bus. Would the sentence "I decided on the car," have a similar double meaning? I made my decision while on the bus.I decided to ride the bus.

We know the dog belongs to the man, and the fountain belongs to the park. It is not clear if the telescope belongs to the man, the woman, or the park.

M12 Contributed by Robert Beezer Statement [40]

In adjacent pairs the words are contradictory or inappropriate. Something cannot be both green and colorless, ideas do not have color, ideas do not sleep, and it is hard to sleep furiously.

M13 Contributed by Robert Beezer Statement [40] Did you assume that the baby and mother are human? Did you assume that the baby is the child of the mother? Did you assume that the mother picked up the baby as an attempt to stop the crying?

M30 Contributed by Robert Beezer Statement [41]

If x, y and z represent the money held by Dan, Diane and Donna, then y = 15 - z and x = 20 - y = 20 - (15 - z) = 5 + z. We can let z take on any value from 0 to 15 without any of the three amounts being negative, since presumably middle-schoolers are too young to assume debt.

Then the total capital held by the three is x + y + z = (5 + z) + (15 - z) + z = 20 + z. So their combined holdings can range anywhere from \$20 (Donna is broke) to \$35 (Donna is flush).

We will have more to say about this situation in Section TSS [112], and specifically Theorem CMVEI [127].

M70 Contributed by Robert Beezer Statement [41]

The equation $x^2 - y^2 = 1$ has a solution set by itself that has the shape of a hyperbola when plotted. Four of the five different second equations have solution sets that are circles when plotted individually (the last is another hyperbola). Where the hyperbola and circles intersect are the solutions to the system of two equations. As the size and location of the circles vary, the number of intersections varies from four to one (in the order given). The last equation is a hyperbola that "opens" in the other direction. Sketching the relevant equations would be instructive, as was discussed in Example STNE [16].

The exact solution sets are (according to the choice of the second equation),

$$\begin{aligned} x^2 + y^2 &= 4: \qquad \left\{ \left(\sqrt{\frac{5}{2}}, \sqrt{\frac{3}{2}} \right), \left(-\sqrt{\frac{5}{2}}, \sqrt{\frac{3}{2}} \right), \left(\sqrt{\frac{5}{2}}, -\sqrt{\frac{3}{2}} \right), \left(-\sqrt{\frac{5}{2}}, -\sqrt{\frac{3}{2}} \right) \right\} \\ x^2 + 2x + y^2 &= 3: \qquad \left\{ (1,0), (-2,\sqrt{3}), (-2,-\sqrt{3}) \right\} \\ x^2 + y^2 &= 1: \qquad \left\{ (1,0), (-1,0) \right\} \\ x^2 - 4x + y^2 &= -3: \qquad \left\{ (1,0) \right\} \\ -x^2 + y^2 &= 1: \qquad \left\{ \right\} \end{aligned}$$

T10 Contributed by Robert Beezer Statement [42] We can say that an integer is **odd** if when it is divided by 2 there is a remainder of 1. So 6 is not odd since $6 = 3 \times 2 + 0$, while 11 is odd since $11 = 5 \times 2 + 1$.

T20 Contributed by Robert Beezer Statement [42] Definition EO [23] is engineered to make Theorem EOPSS [24] true. If we were to allow a zero scalar to multiply an equation then that equation would be transformed to the equation 0 = 0, which is true for any possible values of the variables. Any restrictions on the solution set imposed by the original equation would be lost.

However, in the third operation, it is allowed to choose a zero scalar, multiply an equation by this scalar and add the transformed equation to a second equation (leaving the first unchanged). The result? Nothing. The second equation is the same as it was before. So the theorem is true in this case, the two systems are equivalent. But in practice, this would be a silly thing to actually ever do! We still allow it though, in order to keep our theorem as general as possible.

Notice the location in the proof of Theorem EOPSS [24] where the expression $\frac{1}{\alpha}$ appears — this explains the prohibition on $\alpha = 0$ in the second equation operation.

Section RREF Reduced Row-Echelon Form

After solving a few systems of equations, you will recognize that it doesn't matter so much what we call our variables, as opposed to what numbers act as their coefficients. A system in the variables x_1, x_2, x_3 would behave the same if we changed the names of the variables to a, b, c and kept all the constants the same and in the same places. In this section, we will isolate the key bits of information about a system of equations into something called a matrix, and then use this matrix to systematically solve the equations. Along the way we will obtain one of our most important and useful computational tools.

Subsection MVNSE Matrix and Vector Notation for Systems of Equations

Definition M Matrix

An $m \times n$ matrix is a rectangular layout of numbers from \mathbb{C} having m rows and n columns. We will use upper-case Latin letters from the start of the alphabet (A, B, C, ...) to denote matrices and squared-off brackets to delimit the layout. Many use large parentheses instead of brackets — the distinction is not important. Rows of a matrix will be referenced starting at the top and working down (i.e. row 1 is at the top) and columns will be referenced starting from the left (i.e. column 1 is at the left). For a matrix A, the notation $[A]_{ij}$ will refer to the complex number in row i and column j of A. (This definition contains Notation M.) (This definition contains Notation MC.) \triangle

Be careful with this notation for individual entries, since it is easy to think that $[A]_{ij}$ refers to the *whole* matrix. It does not. It is just a *number*, but is a convenient way to talk about the individual entries simultaneously. This notation will get a heavy workout once we get to Chapter M [425].

Example AM A matrix

$$B = \begin{bmatrix} -1 & 2 & 5 & 3\\ 1 & 0 & -6 & 1\\ -4 & 2 & 2 & -2 \end{bmatrix}$$

is a matrix with m = 3 rows and n = 4 columns. We can say that $[B]_{2,3} = -6$ while $[B]_{3,4} = -2$.

Some mathematical software is very particular about which types of numbers (integers, rationals, reals, complexes) you wish to work with. See: Computation R.SAGE [1583] A calculator or computer language can be a convenient way to perform calculations with

matrices. But first you have to enter the matrix. See: Computation ME.MMA [1566] Computation ME.TI86 [1577] Computation ME.TI83 [1580] Computation ME.SAGE [1584] When we do equation operations on system of equations, the names of the variables really aren't very important. x_1, x_2, x_3 , or a, b, c, or x, y, z, it really doesn't matter. In this subsection we will describe some notation that will make it easier to describe linear systems, solve the systems and describe the solution sets. Here is a list of definitions, laden with notation.

Definition CV

Column Vector

A column vector of size m is an ordered list of m numbers, which is written in order vertically, starting at the top and proceeding to the bottom. At times, we will refer to a column vector as simply a vector. Column vectors will be written in bold, usually with lower case Latin letter from the end of the alphabet such as $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}$. Some books like to write vectors with arrows, such as \vec{u} . Writing by hand, some like to put arrows on top of the symbol, or a tilde underneath the symbol, as in u. To refer to the entry or component that is number i in the list that is the vector \mathbf{v} we write $[\mathbf{v}]_i$.

(This definition contains Notation CV.)

(This definition contains Notation CVC.)

Be careful with this notation. While the symbols $[\mathbf{v}]_i$ might look somewhat substantial, as an object this represents just one component of a vector, which is just a single complex number.

Definition ZCV

Zero Column Vector

The **zero vector** of size m is the column vector of size m where each entry is the number zero,

$$\mathbf{0} = \begin{bmatrix} 0\\0\\0\\\vdots\\0\end{bmatrix}$$

or defined much more compactly, $[\mathbf{0}]_i = 0$ for $1 \le i \le m$. (This definition contains Notation ZCV.)

Definition CM Coefficient Matrix

For a system of linear equations,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$$

the **coefficient matrix** is the $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

Definition VOC Vector of Constants For a system of linear equations,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$$

the **vector of constants** is the column vector of size m

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

Definition SOLV Solution Vector For a system of linear equations,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$$

the **solution vector** is the column vector of size n

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

 \triangle

The solution vector may do double-duty on occasion. It might refer to a list of variable quantities at one point, and subsequently refer to values of those variables that actually form a particular solution to that system.

Definition MRLS

Matrix Representation of a Linear System

If A is the coefficient matrix of a system of linear equations and **b** is the vector of constants, then we will write $\mathcal{LS}(A, \mathbf{b})$ as a shorthand expression for the system of linear equations, which we will refer to as the **matrix representation** of the linear system. (This definition contains Notation MRLS.)

Example NSLE Notation for systems of linear equations

The system of linear equations

 $2x_1 + 4x_2 - 3x_3 + 5x_4 + x_5 = 9$

$$3x_1 + x_2 + x_4 - 3x_5 = 0$$

$$-2x_1 + 7x_2 - 5x_3 + 2x_4 + 2x_5 = -3$$

has coefficient matrix

$$A = \begin{bmatrix} 2 & 4 & -3 & 5 & 1 \\ 3 & 1 & 0 & 1 & -3 \\ -2 & 7 & -5 & 2 & 2 \end{bmatrix}$$

 $\mathbf{b} = \begin{bmatrix} 9\\0\\-3 \end{bmatrix}$

and vector of constants

and so will be referenced as $\mathcal{LS}(A, \mathbf{b})$.

Definition AM Augmented Matrix

Suppose we have a system of m equations in n variables, with coefficient matrix A and vector of constants **b**. Then the **augmented matrix** of the system of equations is the $m \times (n+1)$ matrix whose first n columns are the columns of A and whose last column (number n + 1) is the column vector **b**. This matrix will be written as $[A \mid \mathbf{b}]$. (This definition contains Notation AM.)

The augmented matrix *represents* all the important information in the system of equations, since the names of the variables have been ignored, and the only connection with the variables is the location of their coefficients in the matrix. It is important to realize that the augmented matrix is just that, a matrix, and *not* a system of equations. In particular, the augmented matrix does not have any "solutions," though it will be useful for finding solutions to the system of equations that it is associated with. (Think about your objects, and review Technique L [1614].) However, notice that an augmented matrix always belongs to some system of equations, and vice versa, so it is tempting to try and blur the distinction between the two. Here's a quick example.

Example AMAA Augmented matrix for Archetype A

Archetype A [1642] is the following system of 3 equations in 3 variables.

$$x_1 - x_2 + 2x_3 = 1$$

$$2x_1 + x_2 + x_3 = 8$$

$$x_1 + x_2 = 5$$

Here is its augmented matrix.

$$\begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 8 \\ 1 & 1 & 0 & 5 \end{bmatrix}$$

 \boxtimes

Subsection RO Row Operations

An augmented matrix for a system of equations will save us the tedium of continually writing down the names of the variables as we solve the system. It will also release us from any dependence on the actual names of the variables. We have seen how certain operations we can perform on equations (Definition EO [23]) will preserve their solutions (Theorem EOPSS [24]). The next two definitions and the following theorem carry over these ideas to augmented matrices.

Definition RO Row Operations

The following three operations will transform an $m \times n$ matrix into a different matrix of the same size, and each is known as a **row operation**.

- 1. Swap the locations of two rows.
- 2. Multiply each entry of a single row by a nonzero quantity.
- 3. Multiply each entry of one row by some quantity, and add these values to the entries in the same columns of a second row. Leave the first row the same after this operation, but replace the second row by the new values.

We will use a symbolic shorthand to describe these row operations:

- 1. $R_i \leftrightarrow R_j$: Swap the location of rows *i* and *j*.
- 2. αR_i : Multiply row *i* by the nonzero scalar α .
- 3. $\alpha R_i + R_j$: Multiply row *i* by the scalar α and add to row *j*.

(This definition contains Notation RO.)

Definition REM Row-Equivalent Matrices

Two matrices, A and B, are **row-equivalent** if one can be obtained from the other by a sequence of row operations. \triangle

Example TREM Two row-equivalent matrices

The matrices

$$A = \begin{bmatrix} 2 & -1 & 3 & 4 \\ 5 & 2 & -2 & 3 \\ 1 & 1 & 0 & 6 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 1 & 1 & 0 & 6 \\ 3 & 0 & -2 & -9 \\ 2 & -1 & 3 & 4 \end{bmatrix}$$

are row-equivalent as can be seen from

$$\begin{bmatrix} 2 & -1 & 3 & 4 \\ 5 & 2 & -2 & 3 \\ 1 & 1 & 0 & 6 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & 0 & 6 \\ 5 & 2 & -2 & 3 \\ 2 & -1 & 3 & 4 \end{bmatrix} \xrightarrow{-2R_1 + R_2} \begin{bmatrix} 1 & 1 & 0 & 6 \\ 3 & 0 & -2 & -9 \\ 2 & -1 & 3 & 4 \end{bmatrix}$$

 \wedge

We can also say that any pair of these three matrices are row-equivalent.

Notice that each of the three row operations is reversible (Exercise RREF.T10 [95]), so we do not have to be careful about the distinction between "A is row-equivalent to B" and "B is row-equivalent to A." (Exercise RREF.T11 [95]) The preceding definitions are designed to make the following theorem possible. It says that row-equivalent matrices represent systems of linear equations that have identical solution sets.

Theorem REMES

Row-Equivalent Matrices represent Equivalent Systems

Suppose that A and B are row-equivalent augmented matrices. Then the systems of linear equations that they represent are equivalent systems. \Box

Proof If we perform a single row operation on an augmented matrix, it will have the same effect as if we did the analogous equation operation on the corresponding system of equations. By exactly the same methods as we used in the proof of Theorem EOPSS [24] we can see that each of these row operations will preserve the set of solutions for the corresponding system of equations.

So at this point, our strategy is to begin with a system of equations, represent it by an augmented matrix, perform row operations (which will preserve solutions for the corresponding systems) to get a "simpler" augmented matrix, convert back to a "simpler" system of equations and then solve that system, knowing that its solutions are those of the original system. Here's a rehash of Example US [29] as an exercise in using our new tools.

Example USR

 \boxtimes

Three equations, one solution, reprised

We solve the following system using augmented matrices and row operations. This is the same system of equations solved in Example US [29] using equation operations.

$$x_1 + 2x_2 + 2x_3 = 4$$
$$x_1 + 3x_2 + 3x_3 = 5$$
$$2x_1 + 6x_2 + 5x_3 = 6$$

Form the augmented matrix,

$$A = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 1 & 3 & 3 & 5 \\ 2 & 6 & 5 & 6 \end{bmatrix}$$

and apply row operations,

$$\xrightarrow{-1R_1+R_2} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 2 & 6 & 5 & 6 \end{bmatrix} \xrightarrow{-2R_1+R_3} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & -2 \end{bmatrix}$$

$$\xrightarrow{-2R_2+R_3} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & -4 \end{bmatrix} \xrightarrow{-1R_3} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

So the matrix

$$B = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

is row equivalent to A and by Theorem REMES [61] the system of equations below has the same solution set as the original system of equations.

$$x_1 + 2x_2 + 2x_3 = 4$$
$$x_2 + x_3 = 1$$
$$x_3 = 4$$

Solving this "simpler" system is straightforward and is identical to the process in Example US [29]. \boxtimes

Subsection RREF Reduced Row-Echelon Form

The preceding example amply illustrates the definitions and theorems we have seen so far. But it still leaves two questions unanswered. Exactly what is this "simpler" form for a matrix, and just how do we get it? Here's the answer to the first question, a definition of reduced row-echelon form.

Definition RREF Reduced Row-Echelon Form

A matrix is in **reduced row-echelon form** if it meets all of the following conditions:

- 1. If there is a row where every entry is zero, then this row lies below any other row that contains a nonzero entry.
- 2. The leftmost nonzero entry of a row is equal to 1.
- 3. The leftmost nonzero entry of a row is the only nonzero entry in its column.
- 4. Consider any two different leftmost nonzero entries, one located in row i, column j and the other located in row s, column t. If s > i, then t > j.

A row of only zero entries will be called a **zero row** and the leftmost nonzero entry of a nonzero row will be called a **leading 1**. The number of nonzero rows will be denoted by r.

A column containing a leading 1 will be called a **pivot column**. The set of column indices for all of the pivot columns will be denoted by $D = \{d_1, d_2, d_3, \ldots, d_r\}$ where $d_1 < d_2 < d_3 < \cdots < d_r$, while the columns that are not pivot columns will be denoted as $F = \{f_1, f_2, f_3, \ldots, f_{n-r}\}$ where $f_1 < f_2 < f_3 < \cdots < f_{n-r}$. (This definition contains Notation RREFA.)

The principal feature of reduced row-echelon form is the pattern of leading 1's guaranteed by conditions (2) and (4), reminiscent of a flight of geese, or steps in a staircase, or water cascading down a mountain stream.

There are a number of new terms and notation introduced in this definition, which should make you suspect that this is an important definition. Given all there is to digest here, we will mostly save the use of D and F until Section TSS [112]. However, one important point to make here is that all of these terms and notation apply to a matrix. Sometimes we

will employ these terms and sets for an augmented matrix, and other times it might be a coefficient matrix. So always give some thought to exactly which type of matrix you are analyzing.

Example RREF

A matrix in reduced row-echelon form

The matrix C is in reduced row-echelon form.

This matrix has two zero rows and three leading 1's. So r = 3. Columns 1, 5, and 6 are pivot columns, so $D = \{1, 5, 6\}$ and then $F = \{2, 3, 4, 7, 8\}$.

Example NRREF A matrix not in reduced row-echelon form

The matrix E is not in reduced row-echelon form, as it fails each of the four requirements

once.

$$E = \begin{bmatrix} 1 & 0 & -3 & 0 & 6 & 0 & 7 & -5 & 9 \\ 0 & 0 & 0 & 5 & 0 & 1 & 0 & 3 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

 \boxtimes

Our next theorem has a "constructive" proof. Learn about the meaning of this term in Technique C [1618].

Theorem REMEF Row-Equivalent Matrix in Echelon Form Suppose A is a matrix. Then there is a matrix B so that

- 1. A and B are row-equivalent.
- 2. B is in reduced row-echelon form.

Proof Suppose that A has m rows and n columns. We will describe a process for converting A into B via row operations. This procedure is known as **Gauss–Jordan elimination**. Tracing through this procedure will be easier if you recognize that i refers to a row that

 \square

is being converted, j refers to a column that is being converted, and r keeps track of the number of nonzero rows. Here we go.

1. Set j = 0 and r = 0.

- 2. Increase j by 1. If j now equals n + 1, then stop.
- 3. Examine the entries of A in column j located in rows r + 1 through m. If all of these entries are zero, then go to Step 2.
- 4. Choose a row from rows r + 1 through m with a nonzero entry in column j. Let i denote the index for this row.
- 5. Increase r by 1.
- 6. Use the first row operation to swap rows i and r.
- 7. Use the second row operation to convert the entry in row r and column j to a 1.
- 8. Use the third row operation with row r to convert every other entry of column j to zero.
- 9. Go to Step 2.

The result of this procedure is that the matrix A is converted to a matrix in reduced rowechelon form, which we will refer to as B. We need to now prove this claim by showing that the converted matrix has the requisite properties of Definition RREF [63]. First, the matrix is only converted through row operations (Step 6, Step 7, Step 8), so A and B are row-equivalent (Definition REM [60]).

It is a bit more work to be certain that B is in reduced row-echelon form. We claim that as we begin Step 2, the first j columns of the matrix are in reduced row-echelon form with r nonzero rows. Certainly this is true at the start when j = 0, since the matrix has no columns and so vacuously meets the conditions of Definition RREF [63] with r = 0 nonzero rows.

In Step 2 we increase j by 1 and begin to work with the next column. There are two possible outcomes for Step 3. Suppose that every entry of column j in rows r + 1 through m is zero. Then with no changes we recognize that the first j columns of the matrix has its first r rows still in reduced-row echelon form, with the final m - r rows still all zero.

Suppose instead that the entry in row i of column j is nonzero. Notice that since $r + 1 \leq i \leq m$, we know the first j - 1 entries of this row are all zero. Now, in Step 5 we increase r by 1, and then embark on building a new nonzero row. In Step 6 we swap row r and row i. In the first j columns, the first r - 1 rows remain in reduced row-echelon form after the swap. In Step 7 we multiply row r by a nonzero scalar, creating a 1 in the entry in column j of row i, and not changing any other rows. This new leading 1 is the first nonzero entry in its row, and is located to the right of all the leading 1's in the preceding r - 1 rows. With Step 8 we insure that every entry in the column with this new leading 1 is now zero, as required for reduced row-echelon form. Also, rows r + 1 through m are now all zeros in the first j columns, so we now only have one new nonzero row, consistent with our increase of r by one. Furthermore, since the first j - 1 entries of row r are zero, the employment of the

third row operation does not destroy any of the necessary features of rows 1 through r-1 and rows r+1 through m, in columns 1 through j-1.

So at this stage, the first j columns of the matrix are in reduced row-echelon form. When Step 2 finally increases j to n + 1, then the procedure is completed and the full n columns of the matrix are in reduced row-echelon form, with the value of r correctly recording the number of nonzero rows.

The procedure given in the proof of Theorem REMEF [66] can be more precisely described using a pseudo-code version of a computer program, as follows:

```
input m, n and A

r \leftarrow 0

for j \leftarrow 1 to n

i \leftarrow r+1

while i \leq m and [A]_{ij} = 0

i \leftarrow i+1

if i \neq m+1

r \leftarrow r+1

swap rows i and r of A (row op 1)

scale entry in row r, column j of A to a leading 1 (row op 2)

for k \leftarrow 1 to m, k \neq r

zero out entry in row k, column j of A (row op 3 using row r)

output r and A
```

Notice that as a practical matter the "and" used in the conditional statement of the while

statement should be of the "short-circuit" variety so that the array access that follows is not out-of-bounds.

So now we can put it all together. Begin with a system of linear equations (Definition SLE [18]), and represent the system by its augmented matrix (Definition AM [57]). Use row operations (Definition RO [59]) to convert this matrix into reduced row-echelon form (Definition RREF [63]), using the procedure outlined in the proof of Theorem REMEF [66]. Theorem REMEF [66] also tells us we can always accomplish this, and that the result is row-equivalent (Definition REM [60]) to the original augmented matrix. Since the matrix in reduced-row echelon form has the same solution set, we can analyze the row-reduced version instead of the original matrix, viewing it as the augmented matrix of a different system of equations. The beauty of augmented matrices in reduced row-echelon form is that the solution sets to their corresponding systems can be easily determined, as we will see in the next few examples and in the next section.

We will see through the course that almost every interesting property of a matrix can be discerned by looking at a row-equivalent matrix in reduced row-echelon form. For this reason it is important to know that the matrix B guaranteed to exist by Theorem REMEF [66] is also unique.

Two proof techniques are applicable to the proof. First, head out and read two proof techniques: Technique CD [1623] and Technique U [1625].

Theorem RREFU Reduced Row-Echelon Form is Unique

Suppose that A is an $m \times n$ matrix and that B and C are $m \times n$ matrices that are row-

equivalent to A and in reduced row-echelon form. Then B = C.

Proof We need to begin with no assumptions about any relationships between B and C, other than they are both in reduced row-echelon form, and they are both row-equivalent to A.

If B and C are both row-equivalent to A, then they are row-equivalent to each other. Repeated row operations on a matrix combine the rows with each other using operations that are linear, and are identical in each column. A key observation for this proof is that each individual row of B is linearly related to the rows of C. This relationship is different for each row of B, but once we fix a row, the relationship is the same across columns. More precisely, there are scalars δ_{ik} , $1 \leq i, k \leq m$ such that for any $1 \leq i \leq m$, $1 \leq j \leq n$,

$$[B]_{ij} = \sum_{k=1}^{m} \delta_{ik} [C]_{kj}$$

You should read this as saying that an entry of row *i* of *B* (in column *j*) is a linear function of the entries of all the rows of *C* that are also in column *j*, and the scalars (δ_{ik}) depend on which row of *B* we are considering (the *i* subscript on δ_{ik}), but are the same for every column (no dependence on *j* in δ_{ik}). This idea may be complicated now, but will feel more familiar once we discuss "linear combinations" (Definition LCCV [221]) and moreso when we discuss "row spaces" (Definition RSM [577]). For now, spend some time carefully working Exercise RREF.M40 [94], which is designed to illustrate the origins of this expression. This completes our exploitation of the row-equivalence of *B* and *C*.

We now repeatedly exploit the fact that B and C are in reduced row-echelon form.

П

Recall that a pivot column is all zeros, except a single one. More carefully, if R is a matrix in reduced row-echelon form, and d_{ℓ} is the index of a pivot column, then $[R]_{kd_{\ell}} = 1$ precisely when $k = \ell$ and is otherwise zero. Notice also that any entry of R that is both below the entry in row ℓ and to the left of column d_{ℓ} is also zero (with below and left understood to include equality). In other words, look at examples of matrices in reduced row-echelon form and choose a leading 1 (with a box around it). The rest of the column is also zeros, and the lower left "quadrant" of the matrix that begins here is totally zeros.

Assuming no relationship about the form of B and C, let B have r nonzero rows and denote the pivot columns as $D = \{d_1, d_2, d_3, \ldots, d_r\}$. For C let r' denote the number of nonzero rows and denote the pivot columns as $D' = \{d'_1, d'_2, d'_3, \ldots, d'_{r'}\}$ (Notation RREFA [64]). There are four steps in the proof, and the first three are about showing that B and C have the same number of pivot columns, in the same places. In other words, the "primed" symbols are a necessary fiction.

First Step. Suppose that $d_1 < d'_1$. Then

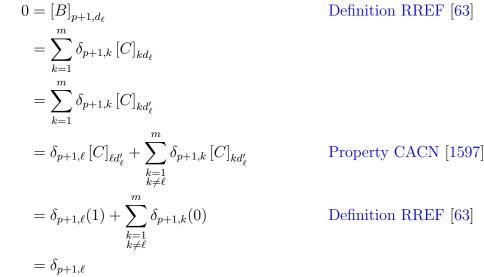
$$1 = [B]_{1d_1}$$
 Definition RREF [63]
$$= \sum_{k=1}^{m} \delta_{1k} [C]_{kd_1}$$

$$= \sum_{k=1}^{m} \delta_{1k} (0)$$

$$= 0$$

The entries of C are all zero since they are left and below of the leading 1 in row 1 and column d'_1 of C. This is a contradiction, so we know that $d_1 \ge d'_1$. By an entirely similar argument, reversing the roles of B and C, we could conclude that $d_1 \leq d'_1$. Together this means that $d_1 = d'_1$.

Second Step. Suppose that we have determined that $d_1 = d'_1, d_2 = d'_2, d_3 = d'_3, \ldots$ $d_p = d'_p$. Let's now show that $d_{p+1} = d'_{p+1}$. Working towards a contradiction, suppose that $d_{p+1} < d'_{n+1}$. For $1 \le \ell \le p$,



Definition RREF [63]

Now,

$$1 = [B]_{p+1,d_{p+1}}$$
 Definition RREF [63]

$$= \sum_{k=1}^{m} \delta_{p+1,k} [C]_{kd_{p+1}}$$
 Property AACN [1597]

$$= \sum_{k=1}^{p} \delta_{p+1,k} [C]_{kd_{p+1}} + \sum_{k=p+1}^{m} \delta_{p+1,k} [C]_{kd_{p+1}}$$
 Property AACN [1597]

$$= \sum_{k=p+1}^{p} (0) [C]_{kd_{p+1}} + \sum_{k=p+1}^{m} \delta_{p+1,k} [C]_{kd_{p+1}}$$

$$= \sum_{k=p+1}^{m} \delta_{p+1,k} [C]_{kd_{p+1}}$$

$$= \sum_{k=p+1}^{m} \delta_{p+1,k} (0)$$
 $d_{p+1} < d'_{p+1}$

$$= 0$$

This contradiction shows that $d_{p+1} \ge d'_{p+1}$. By an entirely similar argument, we could conclude that $d_{p+1} \le d'_{p+1}$, and therefore $d_{p+1} = d'_{p+1}$.

Third Step. Now we establish that r = r'. Suppose that r' < r. By the arguments above,

we know that $d_1 = d'_1, d_2 = d'_2, d_3 = d'_3, \ldots, d_{r'} = d'_{r'}$. For $1 \le \ell \le r' < r$,

$$0 = [B]_{rd_{\ell}} \qquad \text{Definition RREF [63]}$$

$$= \sum_{k=1}^{m} \delta_{rk} [C]_{kd_{\ell}}$$

$$= \sum_{k=1}^{r'} \delta_{rk} [C]_{kd_{\ell}} + \sum_{k=r'+1}^{m} \delta_{rk} [C]_{kd_{\ell}} \qquad \text{Property AACN [1597]}$$

$$= \sum_{k=1}^{r'} \delta_{rk} [C]_{kd_{\ell}} + \sum_{k=r'+1}^{m} \delta_{rk} (0) \qquad \text{Property AACN [1597]}$$

$$= \sum_{k=1}^{r'} \delta_{rk} [C]_{kd_{\ell}}$$

$$= \sum_{k=1}^{r'} \delta_{rk} [C]_{kd_{\ell}}$$

$$= \delta_{r\ell} [C]_{\ell d_{\ell}'} + \sum_{\substack{k=1\\k\neq\ell}}^{r'} \delta_{rk} [C]_{kd_{\ell}} \qquad \text{Property CACN [1597]}$$

$$= \delta_{r\ell}(1) + \sum_{\substack{k=1\\k\neq\ell}}^{r'} \delta_{rk}(0)$$
$$= \delta_{r\ell}$$

Now examine the entries of row r of B,

$$[B]_{rj} = \sum_{k=1}^{m} \delta_{rk} [C]_{kj}$$

= $\sum_{k=1}^{r'} \delta_{rk} [C]_{kj} + \sum_{k=r'+1}^{m} \delta_{rk} [C]_{kj}$
= $\sum_{k=1}^{r'} \delta_{rk} [C]_{kj} + \sum_{k=r'+1}^{m} \delta_{rk} (0)$
= $\sum_{k=1}^{r'} \delta_{rk} [C]_{kj}$
= $\sum_{k=1}^{r'} (0) [C]_{kj}$
= 0

Definition RREF [63]

Property CACN [1597]

Definition RREF [63]

So row r is a totally zero row, contradicting that this should be the bottommost nonzero

row of B. So $r' \geq r$. By an entirely similar argument, reversing the roles of B and C, we would conclude that $r' \leq r$ and therefore r = r'. Thus, combining the first three steps we can say that D = D'. In other words, B and C have the same pivot columns, in the same locations.

Fourth Step. In this final step, we will not argue by contradiction. Our intent is to determine the values of the δ_{ij} . Notice that we can use the values of the d_i interchangeably for B and C. Here we go,

$$1 = [B]_{id_i}$$
 Definition RREF [63]

$$= \sum_{k=1}^{m} \delta_{ik} [C]_{kd_i}$$

$$= \delta_{ii} [C]_{id_i} + \sum_{\substack{k=1 \\ k \neq i}}^{m} \delta_{ik} [C]_{kd_i}$$
 Property CACN [159]

$$= \delta_{ii}(1) + \sum_{\substack{k=1 \\ k \neq i}}^{m} \delta_{ik}(0)$$
 Definition RREF [63]

$$= \delta_{ii}$$

and for $\ell \neq i$



Definition RREF [63]

[1597]

$$= \sum_{k=1}^{m} \delta_{ik} [C]_{kd_{\ell}}$$
$$= \delta_{i\ell} [C]_{\ell d_{\ell}} + \sum_{\substack{k=1\\k \neq \ell}}^{m} \delta_{ik} [C]_{kd_{\ell}}$$
$$= \delta_{i\ell} (1) + \sum_{\substack{k=1\\k \neq \ell}}^{m} \delta_{ik} (0)$$
$$= \delta_{i\ell}$$

Property CACN [1597]

Definition RREF [63]

Finally, having determined the values of the δ_{ij} , we can show that B = C. For $1 \le i \le m$, $1 \le j \le n$,

$$[B]_{ij} = \sum_{k=1}^{m} \delta_{ik} [C]_{kj}$$

= $\delta_{ii} [C]_{ij} + \sum_{\substack{k=1 \ k \neq i}}^{m} \delta_{ik} [C]_{kj}$ Prop
= $(1) [C]_{ij} + \sum_{\substack{k=1 \ k \neq i}}^{m} (0) [C]_{kj}$
= $[C]_{ij}$

Property CACN [1597]

So B and C have equal values in every entry, and so are the same matrix.

We will now run through some examples of using these definitions and theorems to solve some systems of equations. From now on, when we have a matrix in reduced row-echelon form, we will mark the leading 1's with a small box. In your work, you can box 'em, circle 'em or write 'em in a different color — just identify 'em somehow. This device will prove very useful later and is a *very good habit* to start developing right now.

Example SAB Solutions for Archetype B

Let's find the solutions to the following system of equations,

$$-7x_1 - 6x_2 - 12x_3 = -33$$

$$5x_1 + 5x_2 + 7x_3 = 24$$

$$x_1 + 4x_3 = 5$$

First, form the augmented matrix,

$$\begin{bmatrix} -7 & -6 & -12 & -33 \\ 5 & 5 & 7 & 24 \\ 1 & 0 & 4 & 5 \end{bmatrix}$$

and work to reduced row-echelon form, first with j = 1,

$$\xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 5 & 5 & 7 & 24 \\ -7 & -6 & -12 & -33 \end{bmatrix} \xrightarrow{-5R_1 + R_2} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 5 & -13 & -1 \\ -7 & -6 & -12 & -33 \end{bmatrix}$$

$$\xrightarrow{7R_1+R_3} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 5 & -13 & -1 \\ 0 & -6 & 16 & 2 \end{bmatrix}$$

Now, with j = 2,

$$\xrightarrow[]{\frac{1}{5}R_2} \begin{bmatrix} \boxed{1} & 0 & 4 & 5\\ 0 & 1 & \frac{-13}{5} & \frac{-1}{5}\\ 0 & -6 & 16 & 2 \end{bmatrix} \xrightarrow[]{6R_2+R_3} \begin{bmatrix} \boxed{1} & 0 & 4 & 5\\ 0 & \boxed{1} & \frac{-13}{5} & \frac{-1}{5}\\ 0 & 0 & \frac{2}{5} & \frac{4}{5} \end{bmatrix}$$

And finally, with j = 3,

$$\xrightarrow{\frac{5}{2}R_3} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 1 & \frac{-13}{5} & \frac{-1}{5} \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\frac{13}{5}R_3 + R_2} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

$$\xrightarrow{-4R_3 + R_1} \begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

This is now the augmented matrix of a very simple system of equations, namely $x_1 = -3$, $x_2 = 5$, $x_3 = 2$, which has an obvious solution. Furthermore, we can see that this is the *only* solution to this system, so we have determined the entire solution set,

$$S = \left\{ \begin{bmatrix} -3\\5\\2 \end{bmatrix} \right\}$$

You might compare this example with the procedure we used in Example US [29]. \square

Archetypes A and B are meant to contrast each other in many respects. So let's solve Archetype A now.

Example SAA Solutions for Archetype A

Let's find the solutions to the following system of equations,

$$x_1 - x_2 + 2x_3 = 1$$

$$2x_1 + x_2 + x_3 = 8$$

$$x_1 + x_2 = 5$$

First, form the augmented matrix,

$$\begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 8 \\ 1 & 1 & 0 & 5 \end{bmatrix}$$

and work to reduced row-echelon form, first with j = 1,

$$\xrightarrow{-2R_1+R_2} \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 3 & -3 & 6 \\ 1 & 1 & 0 & 5 \end{bmatrix} \xrightarrow{-1R_1+R_3} \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 3 & -3 & 6 \\ 0 & 2 & -2 & 4 \end{bmatrix}$$

Now, with j = 2,

$$\xrightarrow{\frac{1}{3}R_2} \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 2 & -2 & 4 \end{bmatrix} \xrightarrow{1R_2+R_1} \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 2 \\ 0 & 2 & -2 & 4 \end{bmatrix}$$

$$\xrightarrow{-2R_2+R_3} \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The system of equations represented by this augmented matrix needs to be considered a bit differently than that for Archetype B. First, the last row of the matrix is the equation 0 = 0, which is *always* true, so it imposes no restrictions on our possible solutions and therefore we can safely ignore it as we analyze the other two equations. These equations are,

$$x_1 + x_3 = 3$$

 $x_2 - x_3 = 2$

While this system is fairly easy to solve, it also appears to have a multitude of solutions. For example, choose $x_3 = 1$ and see that then $x_1 = 2$ and $x_2 = 3$ will together form a solution. Or choose $x_3 = 0$, and then discover that $x_1 = 3$ and $x_2 = 2$ lead to a solution. Try it yourself: pick *any* value of x_3 you please, and figure out what x_1 and x_2 should be to make the first and second equations (respectively) true. We'll wait while you do that. Because of this behavior, we say that x_3 is a "free" or "independent" variable. But why do we vary x_3 and not some other variable? For now, notice that the third column of the augmented matrix does not have any leading 1's in its column. With this idea, we can rearrange the two equations, solving each for the variable that corresponds to the leading 1 in that row.

$$x_1 = 3 - x_3$$
$$x_2 = 2 + x_3$$

To write the set of solution vectors in set notation, we have

$$S = \left\{ \begin{bmatrix} 3 - x_3 \\ 2 + x_3 \\ x_3 \end{bmatrix} \middle| x_3 \in \mathbb{C} \right\}$$

We'll learn more in the next section about systems with infinitely many solutions and how to express their solution sets. Right now, you might look back at Example IS [31]. \boxtimes

Example SAE Solutions for Archetype E

Let's find the solutions to the following system of equations,

$$2x_1 + x_2 + 7x_3 - 7x_4 = 2$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = 3
x₁ + x₂ + 4x₃ - 5x₄ = 2

First, form the augmented matrix,

$$\begin{bmatrix} 2 & 1 & 7 & -7 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 1 & 1 & 4 & -5 & 2 \end{bmatrix}$$

and work to reduced row-echelon form, first with j = 1,

$$\xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 2 & 1 & 7 & -7 & 2 \end{bmatrix} \xrightarrow{3R_1 + R_2} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ 0 & 7 & 7 & -21 & 9 \\ 2 & 1 & 7 & -7 & 2 \end{bmatrix}$$

$$\xrightarrow{-2R_1 + R_3} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ 0 & 7 & 7 & -21 & 9 \\ 0 & -1 & -1 & 3 & -2 \end{bmatrix}$$

Now, with j = 2,

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ 0 & -1 & -1 & 3 & -2 \\ 0 & 7 & 7 & -21 & 9 \end{bmatrix} \xrightarrow{-1R_2} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 7 & 7 & -21 & 9 \end{bmatrix} \xrightarrow{-1R_2+R_3} \begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 7 & 7 & -21 & 9 \end{bmatrix} \xrightarrow{-7R_2+R_3} \begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & -5 \end{bmatrix}$$

And finally, with j = 4,

$$\xrightarrow{-\frac{1}{5}R_3} \begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{-2R_3+R_2} \begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Let's analyze the equations in the system represented by this augmented matrix. The third equation will read 0 = 1. This is patently false, all the time. No choice of values for our variables will ever make it true. We're done. Since we cannot even make the last equation true, we have no hope of making all of the equations simultaneously true. So this system has no solutions, and its solution set is the empty set, $\emptyset = \{ \}$ (Definition ES [1603]).

Notice that we could have reached this conclusion sooner. After performing the row operation $-7R_2 + R_3$, we can see that the third equation reads 0 = -5, a false statement. Since the system represented by this matrix has no solutions, none of the systems represented has any solutions. However, for this example, we have chosen to bring the matrix fully to reduced row-echelon form for the practice.

These three examples (Example SAB [79], Example SAA [81], Example SAE [83]) illustrate the full range of possibilities for a system of linear equations — no solutions, one solution, or infinitely many solutions. In the next section we'll examine these three scenarios more closely.

Definition RR Row-Reducing

To row-reduce the matrix A means to apply row operations to A and arrive at a row-

equivalent matrix B in reduced row-echelon form.

So the term **row-reduce** is used as a verb. Theorem REMEF [66] tells us that this process will always be successful and Theorem RREFU [70] tells us that the result will be unambiguous. Typically, the analysis of A will proceed by analyzing B and applying theorems whose hypotheses include the row-equivalence of A and B.

After some practice by hand, you will want to use your favorite computing device to do the computations required to bring a matrix to reduced row-echelon form (Exercise RREF.C30 [93]). See: Computation RR.MMA [1567] Computation RR.TI86 [1578] Computation RR.TI83 [1581] Computation RR.SAGE [1586]

Subsection READ Reading Questions

1. Is the matrix below in reduced row-echelon form? Why or why not?

$$\begin{bmatrix} 1 & 5 & 0 & 6 & 8 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

2. Use row operations to convert the matrix below to reduced row-echelon form and report

 \land

the final matrix.

$$\begin{bmatrix} 2 & 1 & 8 \\ -1 & 1 & -1 \\ -2 & 5 & 4 \end{bmatrix}$$

3. Find all the solutions to the system below by using an augmented matrix and row operations. Report your final matrix in reduced row-echelon form and the set of solutions.

$$2x_1 + 3x_2 - x_3 = 0$$

$$x_1 + 2x_2 + x_3 = 3$$

$$x_1 + 3x_2 + 3x_3 = 7$$

Subsection EXC Exercises

C05 Each archetype below is a system of equations. Form the augmented matrix of the system of equations, convert the matrix to reduced row-echelon form by using equation operations and then describe the solution set of the original system of equations.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670] Archetype E [1678] Archetype F [1686] Archetype G [1697] Archetype H [1705] Archetype I [1714] Archetype J [1723] Contributed by Robert Beezer

For problems C10–C19, find all solutions to the system of linear equations. Use your favorite computing device to row-reduce the augmented matrices for the systems, and write the solutions as a set, using correct set notation.

C10

$$2x_1 - 3x_2 + x_3 + 7x_4 = 14$$

$$2x_1 + 8x_2 - 4x_3 + 5x_4 = -1$$

$$x_1 + 3x_2 - 3x_3 = 4$$

$$-5x_1 + 2x_2 + 3x_3 + 4x_4 = -19$$

Contributed by Robert Beezer Solution [97] C11

$$3x_1 + 4x_2 - x_3 + 2x_4 = 6$$

$$x_1 - 2x_2 + 3x_3 + x_4 = 2$$

$$10x_2 - 10x_3 - x_4 = 1$$

Contributed by Robert Beezer Solution [97]

C12

$$2x_1 + 4x_2 + 5x_3 + 7x_4 = -26$$
$$x_1 + 2x_2 + x_3 - x_4 = -4$$
$$-2x_1 - 4x_2 + x_3 + 11x_4 = -10$$

Contributed by Robert Beezer Solution [98] C13

$$x_1 + 2x_2 + 8x_3 - 7x_4 = -2$$

$$3x_1 + 2x_2 + 12x_3 - 5x_4 = 6$$

$$-x_1 + x_2 + x_3 - 5x_4 = -10$$

Contributed by Robert Beezer Solution [99]

C14

$$2x_1 + x_2 + 7x_3 - 2x_4 = 4$$

$$3x_1 - 2x_2 + 11x_4 = 13$$

$$x_1 + x_2 + 5x_3 - 3x_4 = 1$$

Contributed by Robert Beezer Solution [100]

C15

$$2x_1 + 3x_2 - x_3 - 9x_4 = -16$$
$$x_1 + 2x_2 + x_3 = 0$$

$$-x_1 + 2x_2 + 3x_3 + 4x_4 = 8$$

Contributed by Robert Beezer Solution [101] C16

$$2x_1 + 3x_2 + 19x_3 - 4x_4 = 2$$

$$x_1 + 2x_2 + 12x_3 - 3x_4 = 1$$

$$-x_1 + 2x_2 + 8x_3 - 5x_4 = 1$$

Contributed by Robert Beezer Solution [102]

C17

$$-x_1 + 5x_2 = -8$$
$$-2x_1 + 5x_2 + 5x_3 + 2x_4 = 9$$
$$-3x_1 - x_2 + 3x_3 + x_4 = 3$$
$$7x_1 + 6x_2 + 5x_3 + x_4 = 30$$

Contributed by Robert Beezer Solution [102]

C18

$$x_1 + 2x_2 - 4x_3 - x_4 = 32$$

$$x_1 + 3x_2 - 7x_3 - x_5 = 33$$

$$x_1 + 2x_3 - 2x_4 + 3x_5 = 22$$

Contributed by Robert Beezer Solution [103] C19

> $2x_1 + x_2 = 6$ $-x_1 - x_2 = -2$ $3x_1 + 4x_2 = 4$ $3x_1 + 5x_2 = 2$

Contributed by Robert Beezer Solution [104]

For problems C30–C33, row-reduce the matrix without the aid of a calculator, indicating the row operations you are using at each step using the notation of Definition RO [59].

C30

$$\begin{bmatrix} 2 & 1 & 5 & 10 \\ 1 & -3 & -1 & -2 \\ 4 & -2 & 6 & 12 \end{bmatrix}$$

 $\begin{bmatrix} 1 & 2 & -4 \\ -3 & -1 & -3 \\ -2 & 1 & -7 \end{bmatrix}$

Contributed by Robert Beezer

Solution [105]

C31

Contributed	bv	Robert	Beezer
Communutu	Ŋу	1000010	DCCZCI

Solution [106]

C32

$$\begin{bmatrix} 1 & 1 & 1 \\ -4 & -3 & -2 \\ 3 & 2 & 1 \end{bmatrix}$$

Contributed by Robert Beezer

Solution [106]

C33

$$\begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 4 \\ -1 & -2 & 3 & 5 \end{bmatrix}$$

Contributed by Robert Beezer Solution [107]

M40 Consider the two 3×4 matrices below

$$B = \begin{bmatrix} 1 & 3 & -2 & 2 \\ -1 & -2 & -1 & -1 \\ -1 & -5 & 8 & -3 \end{bmatrix} \qquad \qquad C = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 1 & 1 & 4 & 0 \\ -1 & -1 & -4 & 1 \end{bmatrix}$$

(a) Row-reduce each matrix and determine that the reduced row-echelon forms of B and C are identical. From this argue that B and C are row-equivalent.

(b) In the proof of Theorem RREFU [70], we begin by arguing that entries of rowequivalent matrices are related by way of certain scalars and sums. In this example, we would write that entries of B from row i that are in column j are linearly related to the entries of C in column j from all three rows

$$[B]_{ij} = \delta_{i1} [C]_{1j} + \delta_{i2} [C]_{2j} + \delta_{i3} [C]_{3j} \qquad 1 \le j \le 4$$

For each $1 \le i \le 3$ find the corresponding three scalars in this relationship. So your answer will be nine scalars, determined three at a time.

Contributed by Robert Beezer Solution [107]

M45 You keep a number of lizards, mice and peacocks as pets. There are a total of 108 legs and 30 tails in your menagerie. You have twice as many mice as lizards. How many of each creature do you have?

Contributed by Chris Black Solution [109]

M50 A parking lot has 66 vehicles (cars, trucks, motorcycles and bicycles) in it. There are four times as many cars as trucks. The total number of tires (4 per car or truck, 2 per motorcycle or bicycle) is 252. How many cars are there? How many bicycles? Contributed by Robert Beezer Solution [109]

T10 Prove that each of the three row operations (Definition RO [59]) is reversible. More precisely, if the matrix B is obtained from A by application of a single row operation, show that there is a single row operation that will transform B back into A. Contributed by Robert Beezer Solution [110]

T11 Suppose that A, B and C are $m \times n$ matrices. Use the definition of row-equivalence (Definition REM [60]) to prove the following three facts.

- 1. A is row-equivalent to A.
- 2. If A is row-equivalent to B, then B is row-equivalent to A.
- 3. If A is row-equivalent to B, and B is row-equivalent to C, then A is row-equivalent to C.

A relationship that satisfies these three properties is known as an **equivalence relation**, an important idea in the study of various algebras. This is a formal way of saying that a relationship behaves like equality, without requiring the relationship to be as strict as equality itself. We'll see it again in Theorem SER [1033]. Contributed by Robert Beezer

T12 Suppose that *B* is an $m \times n$ matrix in reduced row-echelon form. Build a new, likely smaller, $k \times \ell$ matrix *C* as follows. Keep any collection of *k* adjacent rows, $k \leq m$. From these rows, keep columns 1 through ℓ , $\ell \leq n$. Prove that *C* is in reduced row-echelon form. Contributed by Robert Beezer

T13 Generalize Exercise RREF.T12 [96] by just keeping any k rows, and not requiring the rows to be adjacent. Prove that any such matrix C is in reduced row-echelon form. Contributed by Robert Beezer

Subsection SOL Solutions

C10 Contributed by Robert Beezer Statement [89] The augmented matrix row-reduces to

1	0	0	0	1]	
0	1	0	0	-3	
0	0	1	0	-4	
0	0	0	1	1	

This augmented matrix represents the linear system $x_1 = 1$, $x_2 = -3$, $x_3 = -4$, $x_4 = 1$, which clearly has only one possible solution. We can write this solution set then as

$$S = \left\{ \begin{bmatrix} 1\\ -3\\ -4\\ 1 \end{bmatrix} \right\}$$

C11 Contributed by Robert Beezer Statement [89] The augmented matrix row-reduces to

$$\begin{bmatrix} 1 & 0 & 1 & 4/5 & 0 \\ 0 & 1 & -1 & -1/10 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Row 3 represents the equation 0 = 1, which is patently false, so the original system has no solutions. We can express the solution set as the empty set, $\emptyset = \{\}$.

C12 Contributed by Robert Beezer Statement [89] The augmented matrix row-reduces to

$$\begin{bmatrix} 1 & 2 & 0 & -4 & 2 \\ 0 & 0 & 1 & 3 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In the spirit of Example SAA [81], we can express the infinitely many solutions of this system compactly with set notation. The key is to express certain variables in terms of others. More specifically, each pivot column number is the index of a variable that can be written in terms of the variables whose indices are non-pivot columns. Or saying the same thing: for each iin D, we can find an expression for x_i in terms of the variables without their index in D. Here $D = \{1, 3\}$, so

$$x_1 = 2 - 2x_2 + 4x_4$$
$$x_3 = -6 - 3x_4$$

As a set, we write the solutions precisely as

$$\left\{ \begin{bmatrix} 2 - 2x_2 + 4x_4 \\ x_2 \\ -6 - 3x_4 \\ x_4 \end{bmatrix} \middle| x_2, x_4 \in \mathbb{C} \right\}$$

C13 Contributed by Robert Beezer Statement [90] The augmented matrix of the system of equations is

[1]	2	8	-7	-2]
3	2	12	-5	6
$\lfloor -1 \rfloor$	1	1	-5	$\begin{bmatrix} -2\\ 6\\ -10 \end{bmatrix}$

which row-reduces to

1	0	2	1	0]
0	1	3	-4	0
0	0	0	0	$\begin{array}{c} 0\\ 1\end{array}$

Row 3 represents the equation 0 = 1, which is patently false, so the original system has no solutions. We can express the solution set as the empty set, $\emptyset = \{\}$.

C14 Contributed by Robert Beezer Statement [90]

The augmented matrix of the system of equations is

$$\begin{bmatrix} 2 & 1 & 7 & -2 & 4 \\ 3 & -2 & 0 & 11 & 13 \\ 1 & 1 & 5 & -3 & 1 \end{bmatrix}$$

which row-reduces to

$$\begin{bmatrix} 1 & 0 & 2 & 1 & 3 \\ 0 & 1 & 3 & -4 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In the spirit of Example SAA [81], we can express the infinitely many solutions of this system compactly with set notation. The key is to express certain variables in terms of others. More specifically, each pivot column number is the index of a variable that can be written in terms of the variables whose indices are non-pivot columns. Or saying the same thing: for each iin D, we can find an expression for x_i in terms of the variables without their index in D. Here $D = \{1, 2\}$, so rearranging the equations represented by the two nonzero rows to gain expressions for the variables x_1 and x_2 yields the solution set,

$$S = \left\{ \begin{bmatrix} 3 - 2x_3 - x_4 \\ -2 - 3x_3 + 4x_4 \\ x_3 \\ x_4 \end{bmatrix} \middle| x_3, x_4 \in \mathbb{C} \right\}$$

C15 Contributed by Robert Beezer Statement [90]

The augmented matrix of the system of equations is

$$\begin{bmatrix} 2 & 3 & -1 & -9 & -16 \\ 1 & 2 & 1 & 0 & 0 \\ -1 & 2 & 3 & 4 & 8 \end{bmatrix}$$

which row-reduces to

1	0	0	2	3]
0	1	0	-3	-5
0	0	1	4	7

In the spirit of Example SAA [81], we can express the infinitely many solutions of this system compactly with set notation. The key is to express certain variables in terms of others. More specifically, each pivot column number is the index of a variable that can be written in terms of the variables whose indices are non-pivot columns. Or saying the same thing: for each iin D, we can find an expression for x_i in terms of the variables without their index in D. Here $D = \{1, 2, 3\}$, so rearranging the equations represented by the three nonzero rows to gain expressions for the variables x_1 , x_2 and x_3 yields the solution set,

$$S = \left\{ \begin{bmatrix} 3 - 2x_4 \\ -5 + 3x_4 \\ 7 - 4x_4 \\ x_4 \end{bmatrix} \middle| x_4 \in \mathbb{C} \right\}$$

C16 Contributed by Robert Beezer Statement [91]

The augmented matrix of the system of equations is

$$\begin{bmatrix} 2 & 3 & 19 & -4 & 2 \\ 1 & 2 & 12 & -3 & 1 \\ -1 & 2 & 8 & -5 & 1 \end{bmatrix}$$

which row-reduces to

1	0	2	1	
0	1	5	$-2 \\ 0$	0
0	0	0	0	1

Row 3 represents the equation 0 = 1, which is patently false, so the original system has no solutions. We can express the solution set as the empty set, $\emptyset = \{\}$.

C17 Contributed by Robert Beezer Statement [91] We row-reduce the augmented matrix of the system of equations,

$$\begin{bmatrix} -1 & 5 & 0 & 0 & -8 \\ -2 & 5 & 5 & 2 & 9 \\ -3 & -1 & 3 & 1 & 3 \\ 7 & 6 & 5 & 1 & 30 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

This augmented matrix represents the linear system $x_1 = 3, x_2 = -1, x_3 = 2, x_4 = 5$, which

clearly has only one possible solution. We can write this solution set then as

$$S = \left\{ \begin{bmatrix} 3\\ -1\\ 2\\ 5 \end{bmatrix} \right\}$$

C18 Contributed by Robert Beezer Statement [92] We row-reduce the augmented matrix of the system of equations,

$$\begin{bmatrix} 1 & 2 & -4 & -1 & 0 & 32 \\ 1 & 3 & -7 & 0 & -1 & 33 \\ 1 & 0 & 2 & -2 & 3 & 22 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2 & 0 & 5 & 6 \\ 0 & 1 & -3 & 0 & -2 & 9 \\ 0 & 0 & 0 & 1 & 1 & -8 \end{bmatrix}$$

In the spirit of Example SAA [81], we can express the infinitely many solutions of this system compactly with set notation. The key is to express certain variables in terms of others. More specifically, each pivot column number is the index of a variable that can be written in terms of the variables whose indices are non-pivot columns. Or saying the same thing: for each iin D, we can find an expression for x_i in terms of the variables without their index in D. Here $D = \{1, 2, 4\}$, so

$$\begin{array}{rcl} x_1 + 2x_3 + 5x_5 = 6 & \rightarrow & x_1 = 6 - 2x_3 - 5x_5 \\ x_2 - 3x_3 - 2x_5 = 9 & \rightarrow & x_2 = 9 + 3x_3 + 2x_5 \\ x_4 + x_5 = -8 & \rightarrow & x_4 = -8 - x_5 \end{array}$$

As a set, we write the solutions precisely as

$$S = \left\{ \begin{bmatrix} 6 - 2x_3 - 5x_5\\9 + 3x_3 + 2x_5\\x_3\\-8 - x_5\\x_5 \end{bmatrix} \middle| x_3, x_5 \in \mathbb{C} \right\}$$

C19 Contributed by Robert Beezer Statement [92] We form the augmented matrix of the system,

$$\begin{bmatrix} 2 & 1 & 6 \\ -1 & -1 & -2 \\ 3 & 4 & 4 \\ 3 & 5 & 2 \end{bmatrix}$$

which row-reduces to

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This augmented matrix represents the linear system $x_1 = 4$, $x_2 = -2$, 0 = 0, 0 = 0, which

clearly has only one possible solution. We can write this solution set then as

$$S = \left\{ \begin{bmatrix} 4\\ -2 \end{bmatrix} \right\}$$

C30 Contributed by Robert Beezer Statement [93]

$$\begin{array}{c} \begin{bmatrix} 2 & 1 & 5 & 10 \\ 1 & -3 & -1 & -2 \\ 4 & -2 & 6 & 12 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & -3 & -1 & -2 \\ 2 & 1 & 5 & 10 \\ 4 & -2 & 6 & 12 \end{bmatrix} \\ \xrightarrow{-2R_1 + R_2} \begin{bmatrix} 1 & -3 & -1 & -2 \\ 0 & 7 & 7 & 14 \\ 4 & -2 & 6 & 12 \end{bmatrix} \xrightarrow{-4R_1 + R_3} \begin{bmatrix} 1 & -3 & -1 & -2 \\ 0 & 7 & 7 & 14 \\ 0 & 10 & 10 & 20 \end{bmatrix} \\ \xrightarrow{\frac{1}{7}R_2} \begin{bmatrix} 1 & -3 & -1 & -2 \\ 0 & 1 & 1 & 2 \\ 0 & 10 & 10 & 20 \end{bmatrix} \xrightarrow{3R_2 + R_1} \begin{bmatrix} 1 & 0 & 2 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 10 & 10 & 20 \end{bmatrix} \\ \xrightarrow{-10R_2 + R_3} \begin{bmatrix} 1 & 0 & 2 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

C31 Contributed by Robert Beezer Statement [93]

$$\begin{array}{c} \begin{bmatrix} 1 & 2 & -4 \\ -3 & -1 & -3 \\ -2 & 1 & -7 \end{bmatrix} \xrightarrow{3R_1+R_2} \begin{bmatrix} 1 & 2 & -4 \\ 0 & 5 & -15 \\ -2 & 1 & -7 \end{bmatrix} \\ \xrightarrow{2R_1+R_3} \begin{bmatrix} 1 & 2 & -4 \\ 0 & 5 & -15 \\ 0 & 5 & -15 \end{bmatrix} \xrightarrow{\frac{1}{5}R_2} \begin{bmatrix} 1 & 2 & -4 \\ 0 & 1 & -3 \\ 0 & 5 & -15 \end{bmatrix} \\ \xrightarrow{-2R_2+R_1} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 5 & -15 \end{bmatrix} \xrightarrow{-5R_2+R_3} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

C32 Contributed by Robert Beezer Statement [93] Following the algorithm of Theorem REMEF [66], and working to create pivot columns from left to right, we have

$$\begin{bmatrix} 1 & 1 & 1 \\ -4 & -3 & -2 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{4R_1+R_2} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 3 & 2 & 1 \end{bmatrix} \xrightarrow{-3R_1+R_3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix}$$
$$\xrightarrow{-1R_2+R_1} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & -1 & -2 \end{bmatrix} \xrightarrow{1R_2+R_3} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

C33 Contributed by Robert Beezer Statement [93] Following the algorithm of Theorem REMEF [66], and working to create pivot columns from left to right, we have

$$\begin{bmatrix} 1 & 2 & -1 & -1 \\ 2 & 4 & -1 & 4 \\ -1 & -2 & 3 & 5 \end{bmatrix} \xrightarrow{-2R_1+R_2} \begin{bmatrix} 1 & 2 & -1 & -1 \\ 0 & 0 & 1 & 6 \\ -1 & -2 & 3 & 5 \end{bmatrix}$$

$$\xrightarrow{1R_1+R_3} \begin{bmatrix} 1 & 2 & -1 & -1 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 2 & 4 \end{bmatrix} \xrightarrow{1R_2+R_1} \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 2 & 4 \end{bmatrix}$$

$$\xrightarrow{-2R_2+R_3} \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & -8 \end{bmatrix} \xrightarrow{-\frac{1}{8}R_3} \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{-6R_3+R_2} \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{-5R_3+R_1} \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

M40 Contributed by Robert Beezer Statement [94]

(a) Let R be the common reduced row-echelon form of B and C. A sequence of row operations converts B to R and a second sequence of row operations converts C to R. If we "reverse" the second sequence's order, and reverse each individual row operation (see Exercise RREF.T10 [95]) then we can begin with B, convert to R with the first sequence,

and then convert to C with the reversed sequence. Satisfying Definition REM [60] we can say B and C are row-equivalent matrices.

(b) We will work this carefully for the first row of B and just give the solution for the next two rows. For row 1 of B take i = 1 and we have

$$[B]_{1j} = \delta_{11} [C]_{1j} + \delta_{12} [C]_{2j} + \delta_{13} [C]_{3j} \qquad 1 \le j \le 4$$

If we substitute the four values for j we arrive at four linear equations in the three unknowns $\delta_{11}, \delta_{12}, \delta_{13}$,

$$\begin{array}{lll} (j=1) & [B]_{11} = \delta_{11} \, [C]_{11} + \delta_{12} \, [C]_{21} + \delta_{13} \, [C]_{31} & \Rightarrow & 1 = \delta_{11}(1) + \delta_{12}(1) + \delta_{13}(-1) \\ (j=2) & [B]_{12} = \delta_{11} \, [C]_{12} + \delta_{12} \, [C]_{22} + \delta_{13} \, [C]_{32} & \Rightarrow & 3 = \delta_{11}(2) + \delta_{12}(1) + \delta_{13}(-1) \\ (j=3) & [B]_{13} = \delta_{11} \, [C]_{13} + \delta_{12} \, [C]_{23} + \delta_{13} \, [C]_{33} & \Rightarrow & -2 = \delta_{11}(1) + \delta_{12}(4) + \delta_{13}(-4) \\ (j=4) & [B]_{14} = \delta_{11} \, [C]_{14} + \delta_{12} \, [C]_{24} + \delta_{13} \, [C]_{34} & \Rightarrow & 2 = \delta_{11}(2) + \delta_{12}(0) + \delta_{13}(1) \\ \end{array}$$

We form the augmented matrix of this system and row-reduce to find the solutions,

$$\begin{bmatrix} 1 & 1 & -1 & 1 \\ 2 & 1 & -1 & 3 \\ 1 & 4 & -4 & -2 \\ 2 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So the unique solution is $\delta_{11} = 2$, $\delta_{12} = -3$, $\delta_{13} = -2$. Entirely similar work will lead you to

$$\delta_{21} = -1 \qquad \qquad \delta_{22} = 1 \qquad \qquad \delta_{23} = 1$$

and

$$\delta_{31} = -4$$
 $\delta_{32} = 8$ $\delta_{33} = 5$

M45 Contributed by Chris Black Statement [94] Let l, m, p denote the number of lizards, mice and peacocks. Then the statements from the problem yield the equations:

$$4l + 4m + 2p = 108$$
$$l + m + p = 30$$
$$2l - m = 0$$

We form the augmented matrix for this system and row-reduce

$$\begin{bmatrix} 4 & 4 & 2 & 108 \\ 1 & 1 & 1 & 30 \\ 2 & -1 & 0 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 6 \end{bmatrix}$$

From the row-reduced matrix, we see that we have an equivalent system l = 8, m = 16, and p = 6, which means that you have 8 lizards, 16 mice and 6 peacocks.

M50 Contributed by Robert Beezer Statement [95] Let c, t, m, b denote the number of cars, trucks, motorcycles, and bicycles. Then the statements from the problem yield the equations:

$$c + t + m + b = 66$$

$$c - 4t = 0$$
$$4c + 4t + 2m + 2b = 252$$

We form the augmented matrix for this system and row-reduce

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 66 \\ 1 & -4 & 0 & 0 & 0 \\ 4 & 4 & 2 & 2 & 252 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 48 \\ 0 & 1 & 0 & 0 & 12 \\ 0 & 0 & 1 & 1 & 6 \end{bmatrix}$$

The first row of the matrix represents the equation c = 48, so there are 48 cars. The second row of the matrix represents the equation t = 12, so there are 12 trucks. The third row of the matrix represents the equation m + b = 6 so there are anywhere from 0 to 6 bicycles. We can also say that b is a free variable, but the context of the problem limits it to 7 integer values since you cannot have a negative number of motorcycles.

T10 Contributed by Robert Beezer Statement [95]

If we can reverse each row operation individually, then we can reverse a sequence of row operations. The operations that reverse each operation are listed below, using our shorthand notation. Notice how requiring the scalar α to be non-zero makes the second operation reversible.

$$R_i \leftrightarrow R_j \qquad R_i \leftrightarrow R_j$$
$$\alpha R_i, \ \alpha \neq 0 \qquad \frac{1}{\alpha} R_i$$

$$\alpha R_i + R_j \qquad -\alpha R_i + R_j$$

Section TSS Types of Solution Sets

We will now be more careful about analyzing the reduced row-echelon form derived from the augmented matrix of a system of linear equations. In particular, we will see how to systematically handle the situation when we have infinitely many solutions to a system, and we will prove that every system of linear equations has either zero, one or infinitely many solutions. With these tools, we will be able to solve any system by a well-described method.

Subsection CS Consistent Systems

The computer scientist Donald Knuth said, "Science is what we understand well enough to explain to a computer. Art is everything else." In this section we'll remove solving systems of equations from the realm of art, and into the realm of science. We begin with a definition.

Definition CS

Consistent System

A system of linear equations is **consistent** if it has at least one solution. Otherwise, the system is called **inconsistent**. \triangle

We will want to first recognize when a system is inconsistent or consistent, and in the case of consistent systems we will be able to further refine the types of solutions possible. We will do this by analyzing the reduced row-echelon form of a matrix, using the value of r, and the sets of column indices, D and F, first defined back in Definition RREF [63].

Use of the notation for the elements of D and F can be a bit confusing, since we have subscripted variables that are in turn equal to integers used to index the matrix. However, many questions about matrices and systems of equations can be answered once we know r, D and F. The choice of the letters D and F refer to our upcoming definition of dependent and free variables (Definition IDV [118]). An example will help us begin to get comfortable with this aspect of reduced row-echelon form.

Example RREFN Reduced row-echelon form notation

For the 5×9 matrix

$$B = \begin{bmatrix} 1 & 5 & 0 & 0 & 2 & 8 & 0 & 5 & -1 \\ 0 & 0 & 1 & 0 & 4 & 7 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 3 & 9 & 0 & 3 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

in reduced row-echelon form we have

$$r = 4$$

$$d_1 = 1$$

$$d_2 = 3$$

$$d_3 = 4$$

$$d_4 = 7$$

$$f_1 = 2$$

$$f_2 = 5$$

$$f_3 = 6$$

$$f_4 = 8$$

$$f_5 = 9$$

Notice that the sets

$$D = \{d_1, d_2, d_3, d_4\} = \{1, 3, 4, 7\} \qquad F = \{f_1, f_2, f_3, f_4, f_5\} = \{2, 5, 6, 8, 9\}$$

have nothing in common and together account for all of the columns of B (we say it is a **partition** of the set of column indices).

The number r is the single most important piece of information we can get from the reduced row-echelon form of a matrix. It is defined as the number of nonzero rows, but since each nonzero row has a leading 1, it is also the number of leading 1's present. For each leading 1, we have a pivot column, so r is also the number of pivot columns. Repeating ourselves, r is the number of nonzero rows, the number of leading 1's *and* the number of pivot columns. Across different situations, each of these interpretations of the meaning of r will be useful.

Before proving some theorems about the possibilities for solution sets to systems of equations, let's analyze one particular system with an infinite solution set very carefully as an example. We'll use this technique frequently, and shortly we'll refine it slightly.

Archetypes I and J are both fairly large for doing computations by hand (though not impossibly large). Their properties are very similar, so we will frequently analyze the situation

in Archetype I, and leave you the joy of analyzing Archetype J yourself. So work through Archetype I with the text, by hand and/or with a computer, and then tackle Archetype J yourself (and check your results with those listed). Notice too that the archetypes describing systems of equations each lists the values of r, D and F. Here we go...

Example ISSI

Describing infinite solution sets, Archetype I

Archetype I [1714] is the system of m = 4 equations in n = 7 variables.

$$x_1 + 4x_2 - x_4 + 7x_6 - 9x_7 = 3$$

$$2x_1 + 8x_2 - x_3 + 3x_4 + 9x_5 - 13x_6 + 7x_7 = 9$$

$$2x_3 - 3x_4 - 4x_5 + 12x_6 - 8x_7 = 1$$

$$-x_1 - 4x_2 + 2x_3 + 4x_4 + 8x_5 - 31x_6 + 37x_7 = 4$$

This system has a 4×8 augmented matrix that is row-equivalent to the following matrix (check this!), and which is in reduced row-echelon form (the existence of this matrix is guaranteed by Theorem REMEF [66] and its uniqueness is guaranteed by Theorem RREFU [70]),

$$\begin{bmatrix} 1 & 4 & 0 & 0 & 2 & 1 & -3 & 4 \\ 0 & 0 & 1 & 0 & 1 & -3 & 5 & 2 \\ 0 & 0 & 0 & 1 & 2 & -6 & 6 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So we find that r = 3 and

 $D = \{d_1, d_2, d_3\} = \{1, 3, 4\} \qquad F = \{f_1, f_2, f_3, f_4, f_5\} = \{2, 5, 6, 7, 8\}$

Let *i* denote one of the r = 3 non-zero rows, and then we see that we can solve the corresponding equation represented by this row for the variable x_{d_i} and write it as a linear function of the variables x_{f_1} , x_{f_2} , x_{f_3} , x_{f_4} (notice that $f_5 = 8$ does not reference a variable). We'll do this now, but you can already see how the subscripts upon subscripts takes some getting used to.

$$(i = 1) x_{d_1} = x_1 = 4 - 4x_2 - 2x_5 - x_6 + 3x_7$$

$$(i = 2) x_{d_2} = x_3 = 2 - x_5 + 3x_6 - 5x_7$$

$$(i = 3) x_{d_3} = x_4 = 1 - 2x_5 + 6x_6 - 6x_7$$

Each element of the set $F = \{f_1, f_2, f_3, f_4, f_5\} = \{2, 5, 6, 7, 8\}$ is the index of a variable, except for $f_5 = 8$. We refer to $x_{f_1} = x_2$, $x_{f_2} = x_5$, $x_{f_3} = x_6$ and $x_{f_4} = x_7$ as "free" (or "independent") variables since they are allowed to assume any possible combination of values that we can imagine and we can continue on to build a solution to the system by solving individual equations for the values of the other ("dependent") variables.

Each element of the set $D = \{d_1, d_2, d_3\} = \{1, 3, 4\}$ is the index of a variable. We refer to the variables $x_{d_1} = x_1$, $x_{d_2} = x_3$ and $x_{d_3} = x_4$ as "dependent" variables since they *depend* on the *independent* variables. More precisely, for each possible choice of values for the independent variables we get *exactly one* set of values for the dependent variables that combine to form a solution of the system.

To express the solutions as a set, we write

$$\left(\begin{bmatrix} 4 - 4x_2 - 2x_5 - x_6 + 3x_7 \\ x_2 \\ 2 - x_5 + 3x_6 - 5x_7 \\ 1 - 2x_5 + 6x_6 - 6x_7 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} \right| x_2, x_5, x_6, x_7 \in \mathbb{C} \right\}$$

The condition that $x_2, x_5, x_6, x_7 \in \mathbb{C}$ is how we specify that the variables x_2, x_5, x_6, x_7 are "free" to assume any possible values.

This systematic approach to solving a system of equations will allow us to create a precise description of the solution set for any consistent system once we have found the reduced row-echelon form of the augmented matrix. It will work just as well when the set of free variables is empty and we get just a single solution. And we could program a computer to do it! Now have a whack at Archetype J (Exercise TSS.T10 [136]), mimicking the discussion in this example. We'll still be here when you get back.

Using the reduced row-echelon form of the augmented matrix of a system of equations to determine the nature of the solution set of the system is a very key idea. So let's look at one more example like the last one. But first a definition, and then the example. We mix our metaphors a bit when we call variables free versus dependent. Maybe we should call dependent variables "enslaved"?

Definition IDV Independent and Dependent Variables

Suppose A is the augmented matrix of a consistent system of linear equations and B is a row-equivalent matrix in reduced row-echelon form. Suppose j is the index of a column of B that contains the leading 1 for some row (i.e. column j is a pivot column). Then the variable x_j is **dependent**. A variable that is not dependent is called **independent** or **free**. \triangle

If you studied this definition carefully, you might wonder what to do if the system has n variables and column n + 1 is a pivot column? We will see shortly, by Theorem RCLS [120], that this never happens for a consistent system.

Example FDV Free and dependent variables

Consider the system of five equations in five variables,

$$\begin{aligned} x_1 - x_2 - 2x_3 + x_4 + 11x_5 &= 13\\ x_1 - x_2 + x_3 + x_4 + 5x_5 &= 16\\ 2x_1 - 2x_2 + x_4 + 10x_5 &= 21\\ 2x_1 - 2x_2 - x_3 + 3x_4 + 20x_5 &= 38\\ 2x_1 - 2x_2 + x_3 + x_4 + 8x_5 &= 22 \end{aligned}$$

whose augmented matrix row-reduces to

There are leading 1's in columns 1, 3 and 4, so $D = \{1, 3, 4\}$. From this we know that the variables x_1 , x_3 and x_4 will be dependent variables, and each of the r = 3 nonzero rows of the row-reduced matrix will yield an expression for one of these three variables. The set F is all the remaining column indices, $F = \{2, 5, 6\}$. That $6 \in F$ refers to the column originating from the vector of constants, but the remaining indices in F will correspond to free variables, so x_2 and x_5 (the remaining variables) are our free variables. The resulting three equations that describe our solution set are then,

$(x_{d_1} = x_1)$	$x_1 = 6 + x_2 - 3x_5$
$(x_{d_2} = x_3)$	$x_3 = 1 + 2x_5$
$(x_{d_3} = x_4)$	$x_4 = 9 - 4x_5$

Make sure you understand where these three equations came from, and notice how the location of the leading 1's determined the variables on the left-hand side of each equation.

We can compactly describe the solution set as,

$$S = \left\{ \begin{bmatrix} 6 + x_2 - 3x_5 \\ x_2 \\ 1 + 2x_5 \\ 9 - 4x_5 \\ x_5 \end{bmatrix} \middle| x_2, x_5 \in \mathbb{C} \right\}$$

Notice how we express the freedom for x_2 and x_5 : $x_2, x_5 \in \mathbb{C}$.

Sets are an important part of algebra, and we've seen a few already. Being comfortable with sets is important for understanding and writing proofs. If you haven't already, pay a visit now to Section SET [1602].

We can now use the values of m, n, r, and the independent and dependent variables to categorize the solution sets for linear systems through a sequence of theorems. Through the following sequence of proofs, you will want to consult three proof techniques. See Technique E [1619]. See Technique N [1620]. See Technique CP [1621].

First we have an important theorem that explores the distinction between consistent and inconsistent linear systems.

Theorem RCLS

Recognizing Consistency of a Linear System

Suppose A is the augmented matrix of a system of linear equations with n variables. Suppose also that B is a row-equivalent matrix in reduced row-echelon form with r nonzero rows.

Then the system of equations is inconsistent if and only if the leading 1 of row r is located in column n + 1 of B.

Proof (\Leftarrow) The first half of the proof begins with the assumption that the leading 1 of row r is located in column n+1 of B. Then row r of B begins with n consecutive zeros, finishing with the leading 1. This is a representation of the equation 0 = 1, which is false. Since this equation is false for any collection of values we might choose for the variables, there are no solutions for the system of equations, and it is inconsistent.

 (\Rightarrow) For the second half of the proof, we wish to show that if we assume the system is inconsistent, then the final leading 1 is located in the last column. But instead of proving this directly, we'll form the logically equivalent statement that is the contrapositive, and prove that instead (see Technique CP [1621]). Turning the implication around, and negating each portion, we arrive at the logically equivalent statement: If the leading 1 of row r is not in column n + 1, then the system of equations is consistent.

If the leading 1 for row r is located somewhere in columns 1 through n, then every preceding row's leading 1 is also located in columns 1 through n. In other words, since the last leading 1 is not in the last column, no leading 1 for any row is in the last column, due to the echelon layout of the leading 1's (Definition RREF [63]). We will now construct a solution to the system by setting each dependent variable to the entry of the final column for the row with the corresponding leading 1, and setting each free variable to zero. That sentence is pretty vague, so let's be more precise. Using our notation for the sets D and Ffrom the reduced row-echelon form (Notation RREFA [64]):

$$x_{d_i} = [B]_{i,n+1}, \quad 1 \le i \le r$$
 $x_{f_i} = 0, \quad 1 \le i \le n - r$

These values for the variables make the equations represented by the first r rows of B all true (convince yourself of this). Rows numbered greater than r (if any) are all zero rows, hence represent the equation 0 = 0 and are also all true. We have now identified one solution to the system represented by B, and hence a solution to the system represented by A (Theorem REMES [61]). So we can say the system is consistent (Definition CS [113]).

The beauty of this theorem being an equivalence is that we can unequivocally test to see if a system is consistent or inconsistent by looking at just a single entry of the reduced row-echelon form matrix. We could program a computer to do it!

Notice that for a consistent system the row-reduced augmented matrix has $n + 1 \in F$, so the largest element of F does not refer to a variable. Also, for an inconsistent system, $n+1 \in D$, and it then does not make much sense to discuss whether or not variables are free or dependent since there is no solution. Take a look back at Definition IDV [118] and see why we did not need to consider the possibility of referencing x_{n+1} as a dependent variable.

With the characterization of Theorem RCLS [120], we can explore the relationships between r and n in light of the consistency of a system of equations. First, a situation where we can quickly conclude the inconsistency of a system.

Theorem ISRN

Inconsistent Systems, r and n

Suppose A is the augmented matrix of a system of linear equations in n variables. Suppose also that B is a row-equivalent matrix in reduced row-echelon form with r rows that are not completely zeros. If r = n + 1, then the system of equations is inconsistent.

Proof If r = n + 1, then $D = \{1, 2, 3, ..., n, n + 1\}$ and every column of B contains a

leading 1 and is a pivot column. In particular, the entry of column n + 1 for row r = n + 1 is a leading 1. Theorem RCLS [120] then says that the system is inconsistent.

Do not confuse Theorem ISRN [122] with its converse! Go check out Technique CV [1622] right now.

Next, if a system is consistent, we can distinguish between a unique solution and infinitely many solutions, and furthermore, we recognize that these are the only two possibilities.

Theorem CSRN

Consistent Systems, r and n

Suppose A is the augmented matrix of a *consistent* system of linear equations with n variables. Suppose also that B is a row-equivalent matrix in reduced row-echelon form with r rows that are not zero rows. Then $r \leq n$. If r = n, then the system has a unique solution, and if r < n, then the system has infinitely many solutions.

Proof This theorem contains three implications that we must establish. Notice first that B has n+1 columns, so there can be at most n+1 pivot columns, i.e. $r \le n+1$. If r = n+1, then Theorem ISRN [122] tells us that the system is inconsistent, contrary to our hypothesis. We are left with $r \le n$.

When r = n, we find n - r = 0 free variables (i.e. $F = \{n + 1\}$) and any solution must equal the unique solution given by the first n entries of column n + 1 of B.

When r < n, we have n - r > 0 free variables, corresponding to columns of B without a leading 1, excepting the final column, which also does not contain a leading 1 by Theorem RCLS [120]. By varying the values of the free variables suitably, we can demonstrate infinitely

many solutions.

Subsection FV Free Variables

The next theorem simply states a conclusion from the final paragraph of the previous proof, allowing us to state explicitly the number of free variables for a consistent system.

Theorem FVCS Free Variables for Consistent Systems

Suppose A is the augmented matrix of a *consistent* system of linear equations with n variables. Suppose also that B is a row-equivalent matrix in reduced row-echelon form with r rows that are not completely zeros. Then the solution set can be described with n - r free variables.

Proof See the proof of Theorem CSRN [123].

Example CFV Counting free variables

For each archetype that is a system of equations, the values of n and r are listed. Many also contain a few sample solutions. We can use this information profitably, as illustrated by four examples.

- 1. Archetype A [1642] has n = 3 and r = 2. It can be seen to be consistent by the sample solutions given. Its solution set then has n r = 1 free variables, and therefore will be infinite.
- 2. Archetype B [1652] has n = 3 and r = 3. It can be seen to be consistent by the single sample solution given. Its solution set can then be described with n r = 0 free variables, and therefore will have just the single solution.
- 3. Archetype H [1705] has n = 2 and r = 3. In this case, r = n + 1, so Theorem ISRN [122] says the system is inconsistent. We should not try to apply Theorem FVCS [124] to count free variables, since the theorem only applies to consistent systems. (What would happen if you did?)
- 4. Archetype E [1678] has n = 4 and r = 3. However, by looking at the reduced rowechelon form of the augmented matrix, we find a leading 1 in row 3, column 5. By Theorem RCLS [120] we recognize the system as inconsistent. (Why doesn't this example contradict Theorem ISRN [122]?)

 \boxtimes

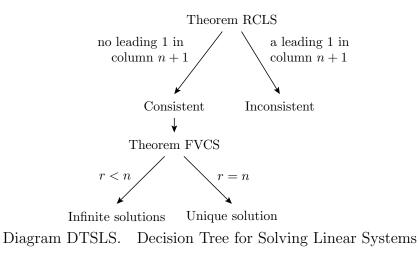
We have accomplished a lot so far, but our main goal has been the following theorem, which is now very simple to prove. The proof is so simple that we ought to call it a corollary, but the result is important enough that it deserves to be called a theorem. (See Technique LC [1634].) Notice that this theorem was presaged first by Example TTS [21] and further foreshadowed by other examples.

Theorem PSSLS Possible Solution Sets for Linear Systems

A system of linear equations has no solutions, a unique solution or infinitely many solutions.

Proof By its definition, a system is either inconsistent or consistent (Definition CS [113]). The first case describes systems with no solutions. For consistent systems, we have the remaining two possibilities as guaranteed by, and described in, Theorem CSRN [123].

Here is a diagram that consolidates several of our theorems from this section, and which is of practical use when you analyze systems of equations.



We have one more theorem to round out our set of tools for determining solution sets to systems of linear equations.

Theorem CMVEI

Consistent, More Variables than Equations, Infinite solutions

Suppose a consistent system of linear equations has m equations in n variables. If n > m, then the system has infinitely many solutions.

Proof Suppose that the augmented matrix of the system of equations is row-equivalent to B, a matrix in reduced row-echelon form with r nonzero rows. Because B has m rows in total, the number that are nonzero rows is less. In other words, $r \leq m$. Follow this with the hypothesis that n > m and we find that the system has a solution set described by at least one free variable because

 $n-r \ge n-m > 0.$

A consistent system with free variables will have an infinite number of solutions, as given by Theorem CSRN [123].

Notice that to use this theorem we need only know that the system is consistent, together with the values of m and n. We do not necessarily have to compute a row-equivalent reduced row-echelon form matrix, even though we discussed such a matrix in the proof. This is the substance of the following example.

Example OSGMD One solution gives many, Archetype D Archetype D is the system of m = 3 equations in n = 4 variables,

$$2x_1 + x_2 + 7x_3 - 7x_4 = 8$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = -12
x₁ + x₂ + 4x₃ - 5x₄ = 4

and the solution $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 1$ can be checked easily by substitution. Having been *handed* this solution, we know the system is consistent. This, together with n > m, allows us to apply Theorem CMVEI [127] and conclude that the system has infinitely many solutions.

These theorems give us the procedures and implications that allow us to completely solve any system of linear equations. The main computational tool is using row operations to convert an augmented matrix into reduced row-echelon form. Here's a broad outline of how we would instruct a computer to solve a system of linear equations.

- 1. Represent a system of linear equations by an augmented matrix (an array is the appropriate data structure in most computer languages).
- 2. Convert the matrix to a row-equivalent matrix in reduced row-echelon form using the procedure from the proof of Theorem REMEF [66].
- 3. Determine r and locate the leading 1 of row r. If it is in column n + 1, output the statement that the system is inconsistent and halt.

- 4. With the leading 1 of row r not in column n + 1, there are two possibilities:
 - (a) r = n and the solution is unique. It can be read off directly from the entries in rows 1 through n of column n + 1.
 - (b) r < n and there are infinitely many solutions. If only a single solution is needed, set all the free variables to zero and read off the dependent variable values from column n + 1, as in the second half of the proof of Theorem RCLS [120]. If the entire solution set is required, figure out some nice compact way to describe it, since your finite computer is not big enough to hold all the solutions (we'll have such a way soon).

The above makes it all sound a bit simpler than it really is. In practice, row operations employ division (usually to get a leading entry of a row to convert to a leading 1) and that will introduce round-off errors. Entries that should be zero sometimes end up being very, very small nonzero entries, or small entries lead to overflow errors when used as divisors. A variety of strategies can be employed to minimize these sorts of errors, and this is one of the main topics in the important subject known as numerical linear algebra.

Solving a linear system is such a fundamental problem in so many areas of mathematics, and its applications, that any computational device worth using for linear algebra will have a built-in routine to do just that. See: Computation LS.MMA [1567] Computation LS.SAGE [1588] In this section we've gained a foolproof procedure for solving any system of linear equations, no matter how many equations or variables. We also have a handful of theorems that allow us to determine partial information about a solution set without

actually constructing the whole set itself. Donald Knuth would be proud.

Subsection READ Reading Questions

- 1. How can we easily recognize when a system of linear equations is inconsistent or not?
- 2. Suppose we have converted the augmented matrix of a system of equations into reduced row-echelon form. How do we then identify the dependent and independent (free) variables?
- 3. What are the possible solution sets for a system of linear equations?

Subsection EXC Exercises

C10 In the spirit of Example ISSI [115], describe the infinite solution set for Archetype J [1723].

Contributed by Robert Beezer

For Exercises C21–C28, find the solution set of the given system of linear equations. Identify the values of n and r, and compare your answers to the results of the theorems of this section.

C21

$$x_1 + 4x_2 + 3x_3 - x_4 = 5$$

$$x_1 - x_2 + x_3 + 2x_4 = 6$$

$$4x_1 + x_2 + 6x_3 + 5x_4 = 9$$

Contributed by Chris Black Solution [138]

C22

$$x_1 - 2x_2 + x_3 - x_4 = 3$$

$$2x_1 - 4x_2 + x_3 + x_4 = 2$$

$$x_1 - 2x_2 - 2x_3 + 3x_4 = 1$$

Contributed by Chris Black Solution [138]

C23

$$x_1 - 2x_2 + x_3 - x_4 = 3$$

$$x_1 + x_2 + x_3 - x_4 = 1$$

$$x_1 + x_3 - x_4 = 2$$

Contributed by Chris Black

Solution [139]

C24

$$x_1 - 2x_2 + x_3 - x_4 = 2$$

$$x_1 + x_2 + x_3 - x_4 = 2$$

$$x_1 + x_3 - x_4 = 2$$

Contributed by Chris Black

Solution [139]

C25

$$x_1 + 2x_2 + 3x_3 = 1$$

$$2x_1 - x_2 + x_3 = 2$$

$$3x_1 + x_2 + x_3 = 4$$

$$x_2 + 2x_3 = 6$$

Contributed by Chris Black Solution [140]

C26

$$x_1 + 2x_2 + 3x_3 = 1$$

$$2x_1 - x_2 + x_3 = 2$$

$$3x_1 + x_2 + x_3 = 4$$

$$5x_2 + 2x_3 = 1$$

Contributed by Chris Black Solution [141]

C27

$$x_1 + 2x_2 + 3x_3 = 0$$

$$2x_1 - x_2 + x_3 = 2$$

$$x_1 - 8x_2 - 7x_3 = 1$$

$$x_2 + x_3 = 0$$

Contributed by Chris Black Solution [141]

C28

$$x_1 + 2x_2 + 3x_3 = 1$$

$$2x_1 - x_2 + x_3 = 2$$

$$x_1 - 8x_2 - 7x_3 = 1$$

$$x_2 + x_3 = 0$$

Contributed by Chris Black Solution [142]

M45 The details for Archetype J [1723] include several sample solutions. Verify that one of these solutions is correct (any one, but just one). Based *only* on this evidence, and especially without doing any row operations, explain how you know this system of linear equations has infinitely many solutions.

Contributed by Robert Beezer Solution [142]

M46 Consider Archetype J [1723], and specifically the row-reduced version of the augmented matrix of the system of equations, denoted as B here, and the values of r, D and F immediately following. Determine the values of the entries

$$[B]_{1,d_1} \quad [B]_{3,d_3} \quad [B]_{1,d_3} \quad [B]_{3,d_1} \quad [B]_{d_1,1} \quad [B]_{d_3,3} \quad [B]_{d_1,3} \quad [B]_{d_3,1} \quad [B]_{1,f_1} \quad [B]_{3,f_1}$$

(See Exercise TSS.M70 [136] for a generalization.) Contributed by Manley Perkel

For Exercises M51–M57 say as much as possible about each system's solution set. Be sure to make it clear which theorems you are using to reach your conclusions. M51 A consistent system of 8 equations in 6 variables. Contributed by Robert Beezer Solution [143]

M52 A consistent system of 6 equations in 8 variables. Contributed by Robert Beezer Solution [143]

M53 A system of 5 equations in 9 variables. Contributed by Robert Beezer Solution [143]

M54 A system with 12 equations in 35 variables. Contributed by Robert Beezer Solution [143]

M56 A system with 6 equations in 12 variables. Contributed by Robert Beezer Solution [143]

M57 A system with 8 equations and 6 variables. The reduced row-echelon form of the augmented matrix of the system has 7 pivot columns. Contributed by Robert Beezer Solution [144]

M60 Without doing any computations, and without examining any solutions, say as much as possible about the form of the solution set for each archetype that is a system of equations.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670] Archetype E [1678] Archetype F [1686] Archetype G [1697] Archetype H [1705] Archetype I [1714] Archetype J [1723] Contributed by Robert Beezer

M70 Suppose that *B* is a matrix in reduced row-echelon form that is equivalent to the augmented matrix of a system of equations with *m* equations in *n* variables. Let *r*, *D* and *F* be as defined in Notation RREFA [64]. What can you conclude, in general, about the following entries?

 $[B]_{1,d_1}$ $[B]_{3,d_3}$ $[B]_{1,d_3}$ $[B]_{3,d_1}$ $[B]_{d_1,1}$ $[B]_{d_3,3}$ $[B]_{d_1,3}$ $[B]_{d_3,1}$ $[B]_{1,f_1}$ $[B]_{3,f_1}$ If you cannot conclude anything about an entry, then say so. (See Exercise TSS.M46 [134] for inspiration.) Contributed by Manley Perkel

T10 An inconsistent system may have r > n. If we try (incorrectly!) to apply Theorem FVCS [124] to such a system, how many free variables would we discover?

Contributed by Robert Beezer Solution [144]

T20 Suppose that *B* is a matrix in reduced row-echelon form that is equivalent to the augmented matrix of a system of equations with *m* equations in *n* variables. Let *r*, *D* and *F* be as defined in Notation RREFA [64]. Prove that $d_k \ge k$ for all $1 \le k \le r$. Then suppose that $r \ge 2$ and $1 \le k < \ell \le r$ and determine what can you conclude, in general, about the following entries.

 $[B]_{k,d_k} [B]_{k,d_{\ell}} [B]_{\ell,d_k} [B]_{d_k,k} [B]_{d_k,\ell} [B]_{d_\ell,k} [B]_{d_\ell,k} [B]_{d_k,f_{\ell}} [B]_{d_\ell,f_k}$

If you cannot conclude anything about an entry, then say so. (See Exercise TSS.M46 [134] and Exercise TSS.M70 [136].) Contributed by Manley Perkel

T40 Suppose that the coefficient matrix of a consistent system of linear equations has two columns that are identical. Prove that the system has infinitely many solutions. Contributed by Robert Beezer Solution [144]

T41 Consider the system of linear equations $\mathcal{LS}(A, \mathbf{b})$, and suppose that every element of the vector of constants \mathbf{b} is a common multiple of the corresponding element of a certain column of A. More precisely, there is a complex number α , and a column index j, such that $[\mathbf{b}]_i = \alpha [A]_{ij}$ for all i. Prove that the system is consistent. Contributed by Robert Beezer Solution [144]

Subsection SOL Solutions

C21 Contributed by Chris Black Statement [131] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 4 & 3 & -1 & 5 \\ 1 & -1 & 1 & 2 & 6 \\ 4 & 1 & 6 & 5 & 9 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 7/5 & 7/5 & 0 \\ 0 & 1 & 2/5 & -3/5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For this system, we have n = 4 and r = 3. However, with a leading 1 in the last column we see that the original system has no solution by Theorem RCLS [120].

C22 Contributed by Chris Black Statement [131] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & -2 & 1 & -1 & 3 \\ 2 & -4 & 1 & 1 & 2 \\ 1 & -2 & -2 & 3 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix}$$

Thus, we see we have an equivalent system for any scalar x_2 :

$$x_1 = 3 + 2x_2$$

$$x_3 = -2$$
$$x_4 = -2.$$

For this system, n = 4 and r = 3. Since it is a consistent system by Theorem RCLS [120], Theorem CSRN [123] guarantees an infinite number of solutions.

C23 Contributed by Chris Black Statement [132] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & -2 & 1 & -1 & 3 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 0 & 1 & -1 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For this system, we have n = 4 and r = 3. However, with a leading 1 in the last column we see that the original system has no solution by Theorem RCLS [120].

C24 Contributed by Chris Black Statement [132] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & -2 & 1 & -1 & 2 \\ 1 & 1 & 1 & -1 & 2 \\ 1 & 0 & 1 & -1 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & -1 & 2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, we see that an equivalent system is

$$x_1 = 2 - x_3 + x_4$$

$$x_2 = 0,$$

and the solution set is $\left\{ \begin{bmatrix} 2 - x_3 + x_4 \\ 0 \\ x_3 \\ x_4 \end{bmatrix} \middle| x_3, x_4 \in \mathbb{C} \right\}.$ For this system, n = 4 and r = 2.

Since it is a consistent system by Theorem RCLS [120], Theorem CSRN [123] guarantees an infinite number of solutions.

C25 Contributed by Chris Black Statement [132] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & -1 & 1 & 2 \\ 3 & 1 & 1 & 4 \\ 0 & 1 & 2 & 6 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since n = 3 and r = 4 = n + 1, Theorem ISRN [122] guarantees that the system is inconsistent. Thus, we see that the given system has no solution.

C26 Contributed by Chris Black Statement [133]

The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & -1 & 1 & 2 \\ 3 & 1 & 1 & 4 \\ 0 & 5 & 2 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 4/3 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & -1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since r = n = 3 and the system is consistent by Theorem RCLS [120], Theorem CSRN [123] guarantees a unique solution, which is

$$x_1 = 4/3$$

 $x_2 = 1/3$
 $x_3 = -1/3.$

C27 Contributed by Chris Black Statement [133] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & -1 & 1 & 2 \\ 1 & -8 & -7 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

For this system, we have n = 3 and r = 3. However, with a leading 1 in the last column we see that the original system has no solution by Theorem RCLS [120].

C28 Contributed by Chris Black Statement [134] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & -1 & 1 & 2 \\ 1 & -8 & -7 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

For this system, n = 3 and r = 2. Since it is a consistent system by Theorem RCLS [120], Theorem CSRN [123] guarantees an infinite number of solutions. An equivalent system is

$$\begin{aligned} x_1 &= 1 - x_3\\ x_2 &= -x_3, \end{aligned}$$

where x_3 is any scalar. So we can express the solution set as

$$\left\{ \left. \begin{bmatrix} 1 - x_3 \\ -x_3 \\ x_3 \end{bmatrix} \right| x_3 \in \mathbb{C} \right\}$$

M45 Contributed by Robert Beezer Statement [134] Demonstrate that the system is consistent by verifying any one of the four sample solutions provided. Then because n = 9 > 6 = m, Theorem CMVEI [127] gives us the conclusion that the system has infinitely many solutions. Notice that we only know the system will have *at* least 9-6=3 free variables, but very well could have more. We do not know know that r=6, only that $r\leq 6$.

M51 Contributed by Robert Beezer Statement [135] Consistent means there is at least one solution (Definition CS [113]). It will have either a unique solution or infinitely many solutions (Theorem PSSLS [126]).

M52 Contributed by Robert Beezer Statement [135]

With 6 rows in the augmented matrix, the row-reduced version will have $r \leq 6$. Since the system is consistent, apply Theorem CSRN [123] to see that $n - r \geq 2$ implies infinitely many solutions.

M53 Contributed by Robert Beezer Statement [135]

The system could be inconsistent. If it is consistent, then because it has more variables than equations Theorem CMVEI [127] implies that there would be infinitely many solutions. So, of all the possibilities in Theorem PSSLS [126], only the case of a unique solution can be ruled out.

M54 Contributed by Robert Beezer Statement [135]

The system could be inconsistent. If it is consistent, then Theorem CMVEI [127] tells us the solution set will be infinite. So we can be certain that there is not a unique solution.

M56 Contributed by Robert Beezer Statement [135]

The system could be inconsistent. If it is consistent, and since 12 > 6, then Theorem CMVEI [127] says we will have infinitely many solutions. So there are two possibilities. Theorem PSSLS [126] allows to state equivalently that a unique solution is an impossibility.

M57 Contributed by Robert Beezer Statement [135]

7 pivot columns implies that there are r = 7 nonzero rows (so row 8 is all zeros in the reduced row-echelon form). Then n + 1 = 6 + 1 = 7 = r and Theorem ISRN [122] allows to conclude that the system is inconsistent.

T10 Contributed by Robert Beezer Statement [136]

Theorem FVCS [124] will indicate a negative number of free variables, but we can say even more. If r > n, then the only possibility is that r = n + 1, and then we compute n - r = n - (n + 1) = -1 free variables.

T40 Contributed by Robert Beezer Statement [137]

Since the system is consistent, we know there is either a unique solution, or infinitely many solutions (Theorem PSSLS [126]). If we perform row operations (Definition RO [59]) on the augmented matrix of the system, the two equal columns of the coefficient matrix will suffer the same fate, and remain equal in the final reduced row-echelon form. Suppose both of these columns are pivot columns (Definition RREF [63]). Then there is single row containing the two leading 1's of the two pivot columns, a violation of reduced row-echelon form (Definition RREF [63]). So at least one of these columns is not a pivot column, and the column index indicates a free variable in the description of the solution set (Definition IDV [118]). With a free variable, we arrive at an infinite solution set (Theorem FVCS [124]).

T41 Contributed by Robert Beezer Statement [137]

The condition about the multiple of the column of constants will allow you to show that the

following values form a solution of the system $\mathcal{LS}(A, \mathbf{b})$,

$$x_1 = 0$$
 $x_2 = 0$... $x_{j-1} = 0$ $x_j = \alpha$ $x_{j+1} = 0$... $x_{n-1} = 0$ $x_n = 0$

With one solution of the system known, we can say the system is consistent (Definition CS [113]).

A more involved proof can be built using Theorem RCLS [120]. Begin by proving that each of the three row operations (Definition RO [59]) will convert the augmented matrix of the system into another matrix where column j is α times the entry of the same row in the last column. In other words, the "column multiple property" is preserved under row operations. These proofs will get successively more involved as you work through the three operations.

Now construct a proof by contradiction (Technique CD [1623]), by supposing that the system is inconsistent. Then the last column of the reduced row-echelon form of the augmented matrix is a pivot column (Theorem RCLS [120]). Then column j must have a zero in the same row as the leading 1 of the final column. But the "column multiple property" implies that there is an α in column j in the same row as the leading 1. So $\alpha = 0$. By hypothesis, then the vector of constants is the zero vector. However, if we began with a final column of zeros, row operations would never have created a leading 1 in the final column. This contradicts the final column being a pivot column, and therefore the system cannot be inconsistent.

Section HSE Homogeneous Systems of Equations

In this section we specialize to systems of linear equations where every equation has a zero as its constant term. Along the way, we will begin to express more and more ideas in the language of matrices and begin a move away from writing out whole systems of equations. The ideas initiated in this section will carry through the remainder of the course.

Subsection SHS Solutions of Homogeneous Systems

As usual, we begin with a definition.

Definition HS Homogeneous System

A system of linear equations, $\mathcal{LS}(A, \mathbf{b})$ is **homogeneous** if the vector of constants is the zero vector, in other words, if $\mathbf{b} = \mathbf{0}$.

Example AHSAC

Archetype C as a homogeneous system

For each archetype that is a system of equations, we have formulated a similar, yet different, homogeneous system of equations by replacing each equation's constant term with a zero. To wit, for Archetype C [1662], we can convert the original system of equations into the homogeneous system,

 $2x_1 - 3x_2 + x_3 - 6x_4 = 0$ $4x_1 + x_2 + 2x_3 + 9x_4 = 0$ $3x_1 + x_2 + x_3 + 8x_4 = 0$

Can you quickly find a solution to this system without row-reducing the augmented matrix? $\hfill \boxtimes$

As you might have discovered by studying Example AHSAC [147], setting each variable to zero will *always* be a solution of a homogeneous system. This is the substance of the following theorem.

Theorem HSC

Homogeneous Systems are Consistent

Suppose that a system of linear equations is homogeneous. Then the system is consistent. \Box

Proof Set each variable of the system to zero. When substituting these values into each equation, the left-hand side evaluates to zero, no matter what the coefficients are. Since a homogeneous system has zero on the right-hand side of each equation as the constant term,

each equation is true. With one demonstrated solution, we can call the system consistent.

Since this solution is so obvious, we now define it as the trivial solution.

Definition TSHSE

Trivial Solution to Homogeneous Systems of Equations

Suppose a homogeneous system of linear equations has n variables. The solution $x_1 = 0$, $x_2 = 0, \ldots, x_n = 0$ (i.e. $\mathbf{x} = \mathbf{0}$) is called the **trivial solution**.

Here are three typical examples, which we will reference throughout this section. Work through the row operations as we bring each to reduced row-echelon form. Also notice what is similar in each example, and what differs.

Example HUSAB

Homogeneous, unique solution, Archetype B

Archetype B can be converted to the homogeneous system,

$$-11x_1 + 2x_2 - 14x_3 = 0$$

$$23x_1 - 6x_2 + 33x_3 = 0$$

$$14x_1 - 2x_2 + 17x_3 = 0$$

whose augmented matrix row-reduces to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

By Theorem HSC [147], the system is consistent, and so the computation n - r = 3 - 3 = 0 means the solution set contains just a single solution. Then, this lone solution must be the trivial solution.

Example HISAA Homogeneous, infinite solutions, Archetype A

Archetype A [1642] can be converted to the homogeneous system,

$$x_1 - x_2 + 2x_3 = 0$$

$$2x_1 + x_2 + x_3 = 0$$

$$x_1 + x_2 = 0$$

whose augmented matrix row-reduces to

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem HSC [147], the system is consistent, and so the computation n - r = 3 - 2 = 1 means the solution set contains one free variable by Theorem FVCS [124], and hence has infinitely many solutions. We can describe this solution set using the free variable x_3 ,

$$S = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \middle| x_1 = -x_3, x_2 = x_3 \right\} = \left\{ \begin{bmatrix} -x_3 \\ x_3 \\ x_3 \end{bmatrix} \middle| x_3 \in \mathbb{C} \right\}$$

Geometrically, these are points in three dimensions that lie on a line through the origin. \square

Example HISAD Homogeneous, infinite solutions, Archetype D

Archetype D [1670] (and identically, Archetype E [1678]) can be converted to the homogeneous system,

$$2x_1 + x_2 + 7x_3 - 7x_4 = 0$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = 0
x₁ + x₂ + 4x₃ - 5x₄ = 0

whose augmented matrix row-reduces to

$$\begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem HSC [147], the system is consistent, and so the computation n - r = 4 - 2 = 2means the solution set contains two free variables by Theorem FVCS [124], and hence has infinitely many solutions. We can describe this solution set using the free variables x_3 and x_4 ,

$$S = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \middle| x_1 = -3x_3 + 2x_4, x_2 = -x_3 + 3x_4 \right\}$$

$$= \left\{ \begin{bmatrix} -3x_3 + 2x_4 \\ -x_3 + 3x_4 \\ x_3 \\ x_4 \end{bmatrix} \middle| x_3, x_4 \in \mathbb{C} \right\}$$

 \boxtimes

After working through these examples, you might perform the same computations for the slightly larger example, Archetype J [1723].

Notice that when we do row operations on the augmented matrix of a homogeneous system of linear equations the last column of the matrix is all zeros. Any one of the three allowable row operations will convert zeros to zeros and thus, the final column of the matrix in reduced row-echelon form will also be all zeros. So in this case, we may be as likely to reference only the coefficient matrix and presume that we remember that the final column begins with zeros, and after any number of row operations is still zero.

Example HISAD [150] suggests the following theorem.

Theorem HMVEI

Homogeneous, More Variables than Equations, Infinite solutions

Suppose that a homogeneous system of linear equations has m equations and n variables with n > m. Then the system has infinitely many solutions.

Proof We are assuming the system is homogeneous, so Theorem HSC [147] says it is consistent. Then the hypothesis that n > m, together with Theorem CMVEI [127], gives

infinitely many solutions.

Example HUSAB [148] and Example HISAA [149] are concerned with homogeneous systems where n = m and expose a fundamental distinction between the two examples. One has a unique solution, while the other has infinitely many. These are exactly the only two possibilities for a homogeneous system and illustrate that each is possible (unlike the case when n > m where Theorem HMVEI [151] tells us that there is only one possibility for a homogeneous system).

Subsection NSM Null Space of a Matrix

The set of solutions to a homogeneous system (which by Theorem HSC [147] is never empty) is of enough interest to warrant its own name. However, we define it as a property of the coefficient matrix, not as a property of some system of equations.

Definition NSM Null Space of a Matrix

The null space of a matrix A, denoted $\mathcal{N}(A)$, is the set of all the vectors that are solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. \wedge

(This definition contains Notation NSM.)

In the Archetypes (Appendix A [1637]) each example that is a system of equations also

has a corresponding homogeneous system of equations listed, and several sample solutions are given. These solutions will be elements of the null space of the coefficient matrix. We'll look at one example.

Example NSEAI Null space elements of Archetype I

The write-up for Archetype I [1714] lists several solutions of the corresponding homogeneous system. Here are two, written as solution vectors. We can say that they are in the null space of the coefficient matrix for the system of equations in Archetype I [1714].

$$\mathbf{x} = \begin{bmatrix} 3\\0\\-5\\-6\\0\\0\\1 \end{bmatrix} \qquad \qquad \mathbf{y} = \begin{bmatrix} -4\\1\\-3\\-2\\1\\1\\1\\1 \end{bmatrix}$$

However, the vector

$$\mathbf{z} = \begin{bmatrix} 1\\0\\0\\0\\0\\0\\2 \end{bmatrix}$$

is not in the null space, since it is not a solution to the homogeneous system. For example, it fails to even make the first equation true. $\hfill \boxtimes$

Here are two (prototypical) examples of the computation of the null space of a matrix.

Example CNS1 Computing a null space, #1 Let's compute the null space of

$$A = \begin{bmatrix} 2 & -1 & 7 & -3 & -8 \\ 1 & 0 & 2 & 4 & 9 \\ 2 & 2 & -2 & -1 & 8 \end{bmatrix}$$

which we write as $\mathcal{N}(A)$. Translating Definition NSM [152], we simply desire to solve the

homogeneous system $\mathcal{LS}(A, \mathbf{0})$. So we row-reduce the augmented matrix to obtain

$$\begin{bmatrix} 1 & 0 & 2 & 0 & 1 & 0 \\ 0 & 1 & -3 & 0 & 4 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 \end{bmatrix}$$

The variables (of the homogeneous system) x_3 and x_5 are free (since columns 1, 2 and 4 are pivot columns), so we arrange the equations represented by the matrix in reduced row-echelon form to

$$x_1 = -2x_3 - x_5 x_2 = 3x_3 - 4x_5 x_4 = -2x_5$$

So we can write the infinite solution set as sets using column vectors,

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} -2x_3 - x_5 \\ 3x_3 - 4x_5 \\ x_3 \\ -2x_5 \\ x_5 \end{bmatrix} \middle| x_3, x_5 \in \mathbb{C} \right\}$$

Example CNS2

 \boxtimes

Computing a null space, #2

Let's compute the null space of

$$C = \begin{bmatrix} -4 & 6 & 1\\ -1 & 4 & 1\\ 5 & 6 & 7\\ 4 & 7 & 1 \end{bmatrix}$$

which we write as $\mathcal{N}(C)$. Translating Definition NSM [152], we simply desire to solve the homogeneous system $\mathcal{LS}(C, \mathbf{0})$. So we row-reduce the augmented matrix to obtain

$\left[1 \right]$	0	0	0
0	1	0	0
0	0	1	0
0	0	0	0

There are no free variables in the homogeneous system represented by the row-reduced matrix, so there is only the trivial solution, the zero vector, $\mathbf{0}$. So we can write the (trivial) solution set as

$$\mathcal{N}(C) = \{\mathbf{0}\} = \left\{ \begin{bmatrix} 0\\0\\0 \end{bmatrix} \right\}$$

 \boxtimes

Subsection READ Reading Questions

- 1. What is *always* true of the solution set for a homogeneous system of equations?
- 2. Suppose a homogeneous system of equations has 13 variables and 8 equations. How many solutions will it have? Why?
- 3. Describe in words (not symbols) the null space of a matrix.

Subsection EXC Exercises

C10 Each Archetype (Appendix A [1637]) that is a system of equations has a corresponding homogeneous system with the same coefficient matrix. Compute the set of solutions for each. Notice that these solution sets are the null spaces of the coefficient matrices.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678] Archetype F [1686] Archetype G [1697]/ Archetype H [1705] Archetype I [1714] and Archetype J [1723] Contributed by Robert Beezer

C20 Archetype K [1734] and Archetype L [1742] are simply 5×5 matrices (i.e. they are not systems of equations). Compute the null space of each matrix. Contributed by Robert Beezer

For Exercises C21-C23, solve the given homogeneous linear system. Compare your results to the results of the corresponding exercise in Section TSS [112].

C21

$$x_1 + 4x_2 + 3x_3 - x_4 = 0$$

$$x_1 - x_2 + x_3 + 2x_4 = 0$$

$$4x_1 + x_2 + 6x_3 + 5x_4 = 0$$

Contributed by Chris Black Solution [164]

C22

$$x_1 - 2x_2 + x_3 - x_4 = 0$$

$$2x_1 - 4x_2 + x_3 + x_4 = 0$$

$$x_1 - 2x_2 - 2x_3 + 3x_4 = 0$$

Contributed by Chris Black Solution [165]

C23

$$x_1 - 2x_2 + x_3 - x_4 = 0$$

$$x_1 + x_2 + x_3 - x_4 = 0$$

$$x_1 + x_3 - x_4 = 0$$

Contributed by Chris Black Solution [165]

For Exercises C25-C27, solve the given homogeneous linear system. Compare your results to the results of the corresponding exercise in Section TSS [112]. C25

 $x_{1} + 2x_{2} + 3x_{3} = 0$ $2x_{1} - x_{2} + x_{3} = 0$ $3x_{1} + x_{2} + x_{3} = 0$ $x_{2} + 2x_{3} = 0$

Contributed by Chris Black Solution [166]

C26

$$x_{1} + 2x_{2} + 3x_{3} = 0$$

$$2x_{1} - x_{2} + x_{3} = 0$$

$$3x_{1} + x_{2} + x_{3} = 0$$

$$5x_{2} + 2x_{3} = 0$$

Contributed by Chris Black Solution [166]

C27

$$x_1 + 2x_2 + 3x_3 = 0$$

$$2x_1 - x_2 + x_3 = 0$$

$$x_1 - 8x_2 - 7x_3 = 0$$

$$x_2 + x_3 = 0$$

Contributed by Chris Black Solution [167] C30 Compute the null space of the matrix A, $\mathcal{N}(A)$.

$$A = \begin{bmatrix} 2 & 4 & 1 & 3 & 8 \\ -1 & -2 & -1 & -1 & 1 \\ 2 & 4 & 0 & -3 & 4 \\ 2 & 4 & -1 & -7 & 4 \end{bmatrix}$$

Contributed by Robert Beezer Solution [167]

C31 Find the null space of the matrix B, $\mathcal{N}(B)$.

$$B = \begin{bmatrix} -6 & 4 & -36 & 6\\ 2 & -1 & 10 & -1\\ -3 & 2 & -18 & 3 \end{bmatrix}$$

Contributed by Robert Beezer Solution [169]

M45 Without doing any computations, and without examining any solutions, say as much as possible about the form of the solution set for corresponding homogeneous system of equations of each archetype that is a system of equations.

```
Archetype A [1642]
Archetype B [1652]
Archetype C [1662]
Archetype D [1670]/Archetype E [1678]
Archetype F [1686]
Archetype G [1697]/Archetype H [1705]
Archetype I [1714]
Archetype J [1723]
Contributed by Robert Beezer
```

For Exercises M50–M52 say as much as possible about each system's solution set. Be sure to make it clear which theorems you are using to reach your conclusions. M50 A homogeneous system of 8 equations in 8 variables. Contributed by Robert Beezer Solution [169]

M51 A homogeneous system of 8 equations in 9 variables. Contributed by Robert Beezer Solution [170]

M52 A homogeneous system of 8 equations in 7 variables.

Contributed by Robert Beezer Solution [170]

T10 Prove or disprove: A system of linear equations is homogeneous if and only if the system has the zero vector as a solution. Contributed by Martin Jackson Solution [170]

T11 Suppose that two systems of linear equations are equivalent. Prove that if the first system is homogeneous, then the second system is homogeneous. Notice that this will allow us to conclude that two equivalent systems are either both homogeneous or both not homogeneous.

Contributed by Robert Beezer Solution [170]

T12 Give an alternate proof of Theorem HSC [147] that uses Theorem RCLS [120]. Contributed by Ivan Kessler

T20 Consider the homogeneous system of linear equations $\mathcal{LS}(A, \mathbf{0})$, and suppose that $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{bmatrix}$ is one solution to the system of equations. Prove that $\mathbf{v} = \begin{bmatrix} 4u_1 \\ 4u_2 \\ 4u_3 \\ \vdots \\ 4u_n \end{bmatrix}$ is also a solution to $\mathcal{LS}(A, \mathbf{0})$.
Contributed by Robert Beezer Solution [171]

Subsection SOL Solutions

C21 Contributed by Chris Black Statement [159] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 4 & 3 & -1 & 0 \\ 1 & -1 & 1 & 2 & 0 \\ 4 & 1 & 6 & 5 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 7/5 & 7/5 & 0 \\ 0 & 1 & 2/5 & -3/5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, we see that the system is consistent (as predicted by Theorem HSC [147]) and has an infinite number of solutions (as predicted by Theorem HMVEI [151]). With suitable choices of x_3 and x_4 , each solution can be written as

$$\begin{bmatrix} -\frac{7}{5}x_3 - \frac{7}{5}x_4\\ -\frac{2}{5}x_3 + \frac{3}{5}x_4\\ x_3\\ x_4 \end{bmatrix}$$

C22 Contributed by Chris Black Statement [159]

The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & -2 & 1 & -1 & 0 \\ 2 & -4 & 1 & 1 & 0 \\ 1 & -2 & -2 & 3 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Thus, we see that the system is consistent (as predicted by Theorem HSC [147]) and has an infinite number of solutions (as predicted by Theorem HMVEI [151]). With a suitable choice of x_2 , each solution can be written as

$$\begin{bmatrix} 2x_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix}$$

C23 Contributed by Chris Black Statement [159] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & -2 & 1 & -1 & 0 \\ 1 & 1 & 1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, we see that the system is consistent (as predicted by Theorem HSC [147]) and has an infinite number of solutions (as predicted by Theorem HMVEI [151]). With suitable choices

of x_3 and x_4 , each solution can be written as

$$\begin{bmatrix} -x_3 + x_4 \\ 0 \\ x_3 \\ x_4 \end{bmatrix}$$

C25 Contributed by Chris Black Statement [160] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & -1 & 1 & 0 \\ 3 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

An homogeneous system is always consistent (Theorem HSC [147]) and with n = r = 3 an application of Theorem FVCS [124] yields zero free variables. Thus the only solution to the given system is the trivial solution, $\mathbf{x} = \mathbf{0}$.

C26 Contributed by Chris Black Statement [160] The augmented matrix for the given linear system and its row-reduced form are:

$$\begin{bmatrix} 1 & 2 & 3 & 0 \\ 2 & -1 & 1 & 0 \\ 3 & 1 & 1 & 0 \\ 0 & 5 & 2 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

An homogeneous system is always consistent (Theorem HSC [147]) and with n = r = 3 an application of Theorem FVCS [124] yields zero free variables. Thus the only solution to the given system is the trivial solution, $\mathbf{x} = \mathbf{0}$.

C27 Contributed by Chris Black Statement [161] The augmented matrix for the given linear system and its row-reduced form are:

[1	2	3	0		$\left[1 \right]$	0	1	0
2	-1	1	0	RREF	0	1	1	0
1	-8	-7	0	\longrightarrow	0	0	0	0
0	1	1	0	$\xrightarrow{\text{RREF}}$	0	0	0	0

An homogeneous system is always consistent (Theorem HSC [147]) and with n = 3, r = 2an application of Theorem FVCS [124] yields one free variable. With a suitable choice of x_3 each solution can be written in the form

$$\begin{bmatrix} -x_3 \\ -x_3 \\ x_3 \end{bmatrix}$$

C30 Contributed by Robert Beezer Statement [161] Definition NSM [152] tells us that the null space of A is the solution set to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. The augmented matrix of this system is

$$\begin{bmatrix} 2 & 4 & 1 & 3 & 8 & 0 \\ -1 & -2 & -1 & -1 & 1 & 0 \\ 2 & 4 & 0 & -3 & 4 & 0 \\ 2 & 4 & -1 & -7 & 4 & 0 \end{bmatrix}$$

To solve the system, we row-reduce the augmented matrix and obtain,

1	2	0	0	5	0
0	0	1	$\begin{array}{c} 0 \\ 0 \\ \hline 1 \\ 0 \end{array}$	-8	0
0	0	0	1	2 0	0
0	0	0	0	0	0

This matrix represents a system with equations having three dependent variables (x_1, x_3, x_4) and two independent variables $(x_2 \text{ and } x_5)$. These equations rearrange to

$$x_1 = -2x_2 - 5x_5 \qquad \qquad x_3 = 8x_5 \qquad \qquad x_4 = -2x_5$$

So we can write the solution set (which is the requested null space) as

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} -2x_2 - 5x_5\\ x_2\\ 8x_5\\ -2x_5\\ x_5 \end{bmatrix} \middle| x_2, x_5 \in \mathbb{C} \right\}$$

C31 Contributed by Robert Beezer Statement [161]

We form the augmented matrix of the homogeneous system $\mathcal{LS}(B, \mathbf{0})$ and row-reduce the matrix,

$$\begin{bmatrix} -6 & 4 & -36 & 6 & 0 \\ 2 & -1 & 10 & -1 & 0 \\ -3 & 2 & -18 & 3 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2 & 1 & 0 \\ 0 & 1 & -6 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We knew ahead of time that this system would be consistent (Theorem HSC [147]), but we can now see there are n - r = 4 - 2 = 2 free variables, namely x_3 and x_4 (Theorem FVCS [124]). Based on this analysis, we can rearrange the equations associated with each nonzero row of the reduced row-echelon form into an expression for the lone dependent variable as a function of the free variables. We arrive at the solution set to the homogeneous system, which is the null space of the matrix by Definition NSM [152],

$$\mathcal{N}(B) = \left\{ \begin{bmatrix} -2x_3 - x_4\\ 6x_3 - 3x_4\\ x_3\\ x_4 \end{bmatrix} \middle| x_3, x_4 \in \mathbb{C} \right\}$$

M50 Contributed by Robert Beezer Statement [162]

Since the system is homogeneous, we know it has the trivial solution (Theorem HSC [147]). We cannot say anymore based on the information provided, except to say that there is either

a unique solution or infinitely many solutions (Theorem PSSLS [126]). See Archetype A [1642] and Archetype B [1652] to understand the possibilities.

M51 Contributed by Robert Beezer Statement [162] Since there are more variables than equations, Theorem HMVEI [151] applies and tells us that the solution set is infinite. From the proof of Theorem HSC [147] we know that the zero vector is one solution.

M52 Contributed by Robert Beezer Statement [162]

By Theorem HSC [147], we know the system is consistent because the zero vector is always a solution of a homogeneous system. There is no more that we can say, since both a unique solution and infinitely many solutions are possibilities.

T10 Contributed by Robert Beezer Statement [163]

This is a true statement. A proof is:

 (\Rightarrow) Suppose we have a homogeneous system $\mathcal{LS}(A, \mathbf{0})$. Then by substituting the scalar zero for each variable, we arrive at true statements for each equation. So the zero vector is a solution. This is the content of Theorem HSC [147].

(\Leftarrow) Suppose now that we have a generic (i.e. not necessarily homogeneous) system of equations, $\mathcal{LS}(A, \mathbf{b})$ that has the zero vector as a solution. Upon substituting this solution into the system, we discover that each component of **b** must also be zero. So $\mathbf{b} = \mathbf{0}$.

T11 Contributed by Robert Beezer Statement [163]

If the first system is homogeneous, then the zero vector is in the solution set of the system. (See the proof of Theorem HSC [147].)

Since the two systems are equivalent, they have equal solutions sets (Definition ESYS [22]). So the zero vector is in the solution set of the second system. By Exercise HSE.T10 [163] the presence of the zero vector in the solution set implies that the system is homogeneous.

So if any one of two equivalent systems is homogeneous, then they both are homogeneous.

T20 Contributed by Robert Beezer Statement [163] Suppose that a single equation from this system (the i-th one) has the form,

$$a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \dots + a_{in}x_n = 0$$

Evaluate the left-hand side of this equation with the components of the proposed solution vector \mathbf{v} ,

$$a_{i1} (4u_1) + a_{i2} (4u_2) + a_{i3} (4u_3) + \dots + a_{in} (4u_n)$$

= $4a_{i1}u_1 + 4a_{i2}u_2 + 4a_{i3}u_3 + \dots + 4a_{in}u_n$ Commutativity
= $4 (a_{i1}u_1 + a_{i2}u_2 + a_{i3}u_3 + \dots + a_{in}u_n)$ Distributivity
= $4(0)$ **u** solution to $\mathcal{LS}(A, \mathbf{0})$
= 0

So \mathbf{v} makes each equation true, and so is a solution to the system.

Notice that this result is not true if we change $\mathcal{LS}(A, \mathbf{0})$ from a homogeneous system to a non-homogeneous system. Can you create an example of a (non-homogeneous) system with a solution **u** such that **v** is not a solution?

Section NM Nonsingular Matrices

In this section we specialize further and consider matrices with equal numbers of rows and columns, which when considered as coefficient matrices lead to systems with equal numbers of equations and variables. We will see in the second half of the course (Chapter D [882], Chapter E [944] Chapter LT [1071], Chapter R [1260]) that these matrices are especially important.

Subsection NM Nonsingular Matrices

Our theorems will now establish connections between systems of equations (homogeneous or otherwise), augmented matrices representing those systems, coefficient matrices, constant vectors, the reduced row-echelon form of matrices (augmented and coefficient) and solution sets. Be very careful in your reading, writing and speaking about systems of equations, matrices and sets of vectors. A system of equations is not a matrix, a matrix is not a solution set, and a solution set is not a system of equations. Now would be a great time to review the discussion about speaking and writing mathematics in Technique L [1614].

Definition SQM Square Matrix

A matrix with m rows and n columns is **square** if m = n. In this case, we say the matrix has **size** n. To emphasize the situation when a matrix is not square, we will call it **rectangular**.

We can now present one of the central definitions of linear algebra.

Definition NM Nonsingular Matrix

Suppose A is a square matrix. Suppose further that the solution set to the homogeneous linear system of equations $\mathcal{LS}(A, \mathbf{0})$ is $\{\mathbf{0}\}$, in other words, the system has *only* the trivial solution. Then we say that A is a **nonsingular** matrix. Otherwise we say A is a **singular** matrix. \triangle

We can investigate whether any square matrix is nonsingular or not, no matter if the matrix is derived somehow from a system of equations or if it is simply a matrix. The definition says that to perform this investigation we must construct a very specific system of equations (homogeneous, with the matrix as the coefficient matrix) and look at its solution set. We will have theorems in this section that connect nonsingular matrices with systems of equations, creating more opportunities for confusion. Convince yourself now of two observations, (1) we can decide nonsingularity for any square matrix, and (2) the determination of nonsingularity involves the solution set for a certain homogeneous system of equations.

 \wedge

Notice that it makes no sense to call a system of equations nonsingular (the term does not apply to a system of equations), nor does it make any sense to call a 5×7 matrix singular (the matrix is not square).

Example S

A singular matrix, Archetype A

Example HISAA [149] shows that the coefficient matrix derived from Archetype A [1642], specifically the 3×3 matrix,

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

is a singular matrix since there are nontrivial solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$.

Example NM A nonsingular matrix, Archetype B

Example HUSAB [148] shows that the coefficient matrix derived from Archetype B [1652], specifically the 3×3 matrix,

$$B = \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix}$$

is a nonsingular matrix since the homogeneous system, $\mathcal{LS}(B, \mathbf{0})$, has only the trivial solu-

tion.

Notice that we will not discuss Example HISAD [150] as being a singular or nonsingular coefficient matrix since the matrix is not square.

The next theorem combines with our main computational technique (row-reducing a matrix) to make it easy to recognize a nonsingular matrix. But first a definition.

Definition IM Identity Matrix The $m \times m$ **identity matrix**, I_m , is defined by

$$[I_m]_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad 1 \le i, \ j \le m$$

(This definition contains Notation IM.)

Example IM An identity matrix The 4 × 4 identity matrix is

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

 \wedge

 \boxtimes

\boxtimes

Notice that an identity matrix is square, and in reduced row-echelon form. So in particular, if we were to arrive at the identity matrix while bringing a matrix to reduced row-echelon form, then it would have all of the diagonal entries circled as leading 1's.

Theorem NMRRI

Nonsingular Matrices Row Reduce to the Identity matrix

Suppose that A is a square matrix and B is a row-equivalent matrix in reduced row-echelon form. Then A is nonsingular if and only if B is the identity matrix. \Box

Proof (\Leftarrow) Suppose *B* is the identity matrix. When the augmented matrix $[A \mid \mathbf{0}]$ is rowreduced, the result is $[B \mid \mathbf{0}] = [I_n \mid \mathbf{0}]$. The number of nonzero rows is equal to the number of variables in the linear system of equations $\mathcal{LS}(A, \mathbf{0})$, so n = r and Theorem FVCS [124] gives n-r = 0 free variables. Thus, the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has just one solution, which must be the trivial solution. This is exactly the definition of a nonsingular matrix (Definition NM [173]).

(⇒) If A is nonsingular, then the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution, and has no free variables in the description of the solution set. The homogeneous system is consistent (Theorem HSC [147]) so Theorem FVCS [124] applies and tells us there are n-rfree variables. Thus, n-r = 0, and so n = r. So B has n pivot columns among its total of n columns. This is enough to force B to be the $n \times n$ identity matrix I_n (see Exercise NM.T12 [187]).

Notice that since this theorem is an equivalence it will always allow us to determine if a

matrix is either nonsingular or singular. Here are two examples of this, continuing our study of Archetype A and Archetype B.

Example SRR Singular matrix, row-reduced

The coefficient matrix for Archetype A [1642] is

$$A = \begin{bmatrix} 1 & -1 & 2\\ 2 & 1 & 1\\ 1 & 1 & 0 \end{bmatrix}$$

which when row-reduced becomes the row-equivalent matrix

$$B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since this matrix is not the 3×3 identity matrix, Theorem NMRRI [176] tells us that A is a singular matrix.

Example NSR Nonsingular matrix, row-reduced The coefficient matrix for Archetype B [1652] is

$$A = \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix}$$

which when row-reduced becomes the row-equivalent matrix

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since this matrix is the 3×3 identity matrix, Theorem NMRRI [176] tells us that A is a nonsingular matrix.

Subsection NSNM Null Space of a Nonsingular Matrix

Nonsingular matrices and their null spaces are intimately related, as the next two examples illustrate.

Example NSS Null space of a singular matrix Given the coefficient matrix from Archetype A [1642],

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

the null space is the set of solutions to the homogeneous system of equations $\mathcal{LS}(A, \mathbf{0})$ has a solution set and null space constructed in Example HISAA [149] as

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} -x_3 \\ x_3 \\ x_3 \end{bmatrix} \middle| x_3 \in \mathbb{C} \right\}$$

$\mathbf{E}\mathbf{x}\mathbf{a}\mathbf{m}\mathbf{p}$	le NSN	Μ			
Null sp	ace of a	nonsi	ngular n	natrix	c
a	n m	•	· • c		

Given the coefficient matrix from Archetype B [1652],

$$A = \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix}$$

the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has a solution set constructed in Example HUSAB [148] that contains only the trivial solution, so the null space has only a single element,

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} 0\\0\\0 \end{bmatrix} \right\}$$

These two examples illustrate the next theorem, which is another equivalence.

 \boxtimes

 \boxtimes

Theorem NMTNS Nonsingular Matrices have Trivial Null Spaces

Suppose that A is a square matrix. Then A is nonsingular if and only if the null space of A, $\mathcal{N}(A)$, contains only the zero vector, i.e. $\mathcal{N}(A) = \{\mathbf{0}\}$.

Proof The null space of a square *matrix*, A, is equal to the set of solutions to the homogeneous *system*, $\mathcal{LS}(A, \mathbf{0})$. A *matrix* is nonsingular if and only if the set of solutions to the homogeneous *system*, $\mathcal{LS}(A, \mathbf{0})$, has only a trivial solution. These two observations may be chained together to construct the two proofs necessary for each half of this theorem.

The next theorem pulls a lot of big ideas together. Theorem NMUS [180] tells us that we can learn much about solutions to a system of linear equations with a square coefficient matrix by just examining a similar homogeneous system.

Theorem NMUS

Nonsingular Matrices and Unique Solutions

Suppose that A is a square matrix. A is a nonsingular matrix if and only if the system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every choice of the constant vector \mathbf{b} .

Proof (\Leftarrow) The hypothesis for this half of the proof is that the system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for *every* choice of the constant vector \mathbf{b} . We will make a very specific choice for \mathbf{b} : $\mathbf{b} = \mathbf{0}$. Then we know that the system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution. But this is precisely the definition of what it means for A to be nonsingular (Definition NM [173]). That almost seems too easy! Notice that we have not used the full power of our hypothesis, but there is nothing that says we must use a hypothesis to its fullest.

 (\Rightarrow) We assume that A is nonsingular of size $n \times n$, so we know there is a sequence of row operations that will convert A into the identity matrix I_n (Theorem NMRRI [176]). Form the augmented matrix $A' = [A \mid \mathbf{b}]$ and apply this same sequence of row operations to A'. The result will be the matrix $B' = [I_n \mid \mathbf{c}]$, which is in reduced row-echelon form with r = n. Then the augmented matrix B' represents the (extremely simple) system of equations $x_i = [\mathbf{c}]_i$, $1 \le i \le n$. The vector \mathbf{c} is clearly a solution, so the system is consistent (Definition CS [113]). With a consistent system, we use Theorem FVCS [124] to count free variables. We find that there are n - r = n - n = 0 free variables, and so we therefore know that the solution is unique. (This half of the proof was suggested by Asa Scherer.)

This theorem helps to explain part of our interest in nonsingular matrices. If a matrix is nonsingular, then no matter what vector of constants we pair it with, using the matrix as the coefficient matrix will *always* yield a linear system of equations with a solution, and the solution is unique. To determine if a matrix has this property (non-singularity) it is enough to just solve one linear system, the homogeneous system with the matrix as coefficient matrix and the zero vector as the vector of constants (or any other vector of constants, see Exercise MM.T10 [493]).

Formulating the negation of the second part of this theorem is a good exercise. A singular matrix has the property that for *some* value of the vector \mathbf{b} , the system $\mathcal{LS}(A, \mathbf{b})$ does not have a unique solution (which means that it has no solution or infinitely many solutions). We will be able to say more about this case later (see the discussion following Theorem PSPHS [256]). Square matrices that are nonsingular have a long list of interesting properties, which

we will start to catalog in the following, recurring, theorem. Of course, singular matrices will then have all of the opposite properties. The following theorem is a list of equivalences. We want to understand just what is involved with understanding and proving a theorem that says several conditions are equivalent. So have a look at Technique ME [1626] before studying the first in this series of theorems.

Theorem NME1

Nonsingular Matrix Equivalences, Round 1

Suppose that A is a square matrix. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of **b**.

Proof That A is nonsingular is equivalent to each of the subsequent statements by, in turn, Theorem NMRRI [176], Theorem NMTNS [179] and Theorem NMUS [180]. So the statement of this theorem is just a convenient way to organize all these results.

Finally, you may have wondered why we refer to a matrix as *nonsingular* when it creates systems of equations with *single* solutions (Theorem NMUS [180])! I've wondered the same thing. We'll have an opportunity to address this when we get to Theorem SMZD [935]. Can you wait that long?

Subsection READ Reading Questions

- 1. What is the definition of a nonsingular matrix?
- 2. What is the easiest way to recognize if a square matrix is nonsingular or not?
- 3. Suppose we have a system of equations and its coefficient matrix is nonsingular. What can you say about the solution set for this system?

Subsection EXC Exercises . . .

In Exercises C30–C33 determine if the matrix is nonsingular or singular. Give reasons for your answer.

C30

$\left[-3\right]$	1	2	8]
2	0	3	4
1	2	7	-4
5	-1	2	0

Contributed by Robert Beezer Solution [188]

C31

2	3	1	4]
1	1	1	$ \begin{array}{c} 4 \\ 0 \\ 5 \\ 3 \end{array} $
-1	2	3	5
1	2	1	3

Contributed by Robert Beezer Solution [188]

C32

$$\begin{bmatrix} 9 & 3 & 2 & 4 \\ 5 & -6 & 1 & 3 \\ 4 & 1 & 3 & -5 \end{bmatrix}$$

Contributed by Robert Beezer

Solution [189]

C33

[-1]	2	0	3]
1	-3	-2	4
-2	0	4	3
-3	1	-2	3

Contributed by Robert Beezer Solution [189]

C40 Each of the archetypes below is a system of equations with a square coefficient matrix, or is itself a square matrix. Determine if these matrices are nonsingular, or singular. Comment on the null space of each matrix.

Archetype A [1642] Archetype B [1652] Archetype F [1686] Archetype K [1734] Archetype L [1742] Contributed by Robert Beezer C50 Find the null space of the matrix E below.

$$E = \begin{bmatrix} 2 & 1 & -1 & -9 \\ 2 & 2 & -6 & -6 \\ 1 & 2 & -8 & 0 \\ -1 & 2 & -12 & 12 \end{bmatrix}$$

Contributed by Robert Beezer Solution [189]

M30 Let A be the coefficient matrix of the system of equations below. Is A nonsingular or singular? Explain what you could infer about the solution set for the system based only on what you have learned about A being singular or nonsingular.

$$-x_1 + 5x_2 = -8$$

$$-2x_1 + 5x_2 + 5x_3 + 2x_4 = 9$$

$$-3x_1 - x_2 + 3x_3 + x_4 = 3$$

$$7x_1 + 6x_2 + 5x_3 + x_4 = 30$$

Contributed by Robert Beezer Solution [190]

For Exercises M51–M52 say **as much as possible** about each system's solution set. Be sure to make it clear which theorems you are using to reach your conclusions. **M51** 6 equations in 6 variables, singular coefficient matrix. Contributed by Robert Beezer Solution [190] M52 A system with a nonsingular coefficient matrix, not homogeneous. Contributed by Robert Beezer Solution [191]

T10 Suppose that A is a square matrix, and B is a matrix in reduced row-echelon form that is row-equivalent to A. Prove that if A is singular, then the last row of B is a zero row. Contributed by Robert Beezer Solution [191]

T12 Suppose that A is a square matrix. Using the definition of reduced row-echelon form (Definition RREF [63]) carefully, give a proof of the following equivalence: Every column of A is a pivot column if and only if A is the identity matrix (Definition IM [175]). Contributed by Robert Beezer

T30 Suppose that A is a nonsingular matrix and A is row-equivalent to the matrix B. Prove that B is nonsingular. Contributed by Robert Beezer Solution [191]

T90 Provide an alternative for the second half of the proof of Theorem NMUS [180], without appealing to properties of the reduced row-echelon form of the coefficient matrix. In other words, prove that if A is nonsingular, then $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every choice of the constant vector \mathbf{b} . Construct this proof without using Theorem REMEF [66] or Theorem RREFU [70].

Contributed by Robert Beezer Solution [192]

Subsection SOL Solutions

C30 Contributed by Robert Beezer Statement [184] The matrix row-reduces to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

which is the 4×4 identity matrix. By Theorem NMRRI [176] the original matrix must be nonsingular.

C31 Contributed by Robert Beezer Statement [184] Row-reducing the matrix yields,

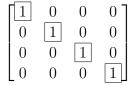
$$\begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since this is not the 4×4 identity matrix, Theorem NMRRI [176] tells us the matrix is singular.

C32 Contributed by Robert Beezer Statement [185]

The matrix is not square, so neither term is applicable. See Definition NM [173], which is stated for just square matrices.

C33 Contributed by Robert Beezer Statement [185] Theorem NMRRI [176] tells us we can answer this question by simply row-reducing the matrix. Doing this we obtain,



Since the reduced row-echelon form of the matrix is the 4×4 identity matrix I_4 , we know that B is nonsingular.

C50 Contributed by Robert Beezer Statement [186] We form the augmented matrix of the homogeneous system $\mathcal{LS}(E, \mathbf{0})$ and row-reduce the matrix,

We knew ahead of time that this system would be consistent (Theorem HSC [147]), but we can now see there are n - r = 4 - 2 = 2 free variables, namely x_3 and x_4 since $F = \{3, 4, 5\}$

(Theorem FVCS [124]). Based on this analysis, we can rearrange the equations associated with each nonzero row of the reduced row-echelon form into an expression for the lone dependent variable as a function of the free variables. We arrive at the solution set to this homogeneous system, which is the null space of the matrix by Definition NSM [152],

$$\mathcal{N}(E) = \left\{ \begin{bmatrix} -2x_3 + 6x_4\\5x_3 - 3x_4\\x_3\\x_4 \end{bmatrix} \middle| x_3, x_4 \in \mathbb{C} \right\}$$

M30 Contributed by Robert Beezer Statement [186] We row-reduce the coefficient matrix of the system of equations,

$$\begin{bmatrix} -1 & 5 & 0 & 0 \\ -2 & 5 & 5 & 2 \\ -3 & -1 & 3 & 1 \\ 7 & 6 & 5 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since the row-reduced version of the coefficient matrix is the 4×4 identity matrix, I_4 (Definition IM [175] byTheorem NMRRI [176], we know the coefficient matrix is nonsingular. According to Theorem NMUS [180] we know that the system is guaranteed to have a unique solution, based only on the extra information that the coefficient matrix is nonsingular.

M51 Contributed by Robert Beezer Statement [186] Theorem NMRRI [176] tells us that the coefficient matrix will not row-reduce to the identity matrix. So if we were to row-reduce the augmented matrix of this system of equations, we would not get a unique solution. So by Theorem PSSLS [126] the remaining possibilities are no solutions, or infinitely many.

M52 Contributed by Robert Beezer Statement [187]

Any system with a nonsingular coefficient matrix will have a unique solution by Theorem NMUS [180]. If the system is not homogeneous, the solution cannot be the zero vector (Exercise HSE.T10 [163]).

T10 Contributed by Robert Beezer Statement [187]

Let *n* denote the size of the square matrix *A*. By Theorem NMRRI [176] the hypothesis that *A* is singular implies that *B* is not the identity matrix I_n . If *B* has *n* pivot columns, then it would have to be I_n , so *B* must have fewer than *n* pivot columns. But the number of nonzero rows in *B* (*r*) is equal to the number of pivot columns as well. So the *n* rows of *B* have fewer than *n* nonzero rows, and *B* must contain at least one zero row. By Definition RREF [63], this row must be at the bottom of *B*.

A proof can also be formulated by first forming the contrapositive of the statement (Technique CP [1621]) and proving this statement.

T30 Contributed by Robert Beezer Statement [187]

Since A and B are row-equivalent matrices, consideration of the three row operations (Definition RO [59]) will show that the augmented matrices, $[A \mid \mathbf{0}]$ and $[B \mid \mathbf{0}]$, are also rowequivalent matrices. This says that the two homogeneous systems, $\mathcal{LS}(A, \mathbf{0})$ and $\mathcal{LS}(B, \mathbf{0})$ are equivalent systems. $\mathcal{LS}(A, \mathbf{0})$ has only the zero vector as a solution (Definition NM [173]), thus $\mathcal{LS}(B, \mathbf{0})$ has only the zero vector as a solution. Finally, by Definition NM [173], we see that B is nonsingular.

Form a similar theorem replacing "nonsingular" by "singular" in both the hypothesis and the conclusion. Prove this new theorem with an approach just like the one above, and/or employ the result about nonsingular matrices in a proof by contradiction.

T90 Contributed by Robert Beezer Statement [187]

We assume A is nonsingular, and try to solve the system $\mathcal{LS}(A, \mathbf{b})$ without making any assumptions about **b**. To do this we will begin by constructing a new homogeneous linear system of equations that looks very much like the original. Suppose A has size n (why must it be square?) and write the original system as,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$$

$$\vdots$$

$$(*)$$

$$a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \dots + a_{nn}x_n = b_n$$

Form the new, homogeneous system in n equations with n + 1 variables, by adding a new variable y, whose coefficients are the negatives of the constant terms,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n - b_1y = 0$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n - b_2y = 0$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n - b_3y = 0$$

$$\vdots$$

$$a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \dots + a_{nn}x_n - b_ny = 0$$

(**)

Since this is a homogeneous system with more variables than equations (m = n + 1 > n), Theorem HMVEI [151] says that the system has infinitely many solutions. We will choose one of these solutions, *any* one of these solutions, so long as it is *not* the trivial solution. Write this solution as

$$x_1 = c_1$$
 $x_2 = c_2$ $x_3 = c_3$... $x_n = c_n$ $y = c_{n+1}$

We know that at least one value of the c_i is nonzero, but we will now show that in particular $c_{n+1} \neq 0$. We do this using a proof by contradiction (Technique CD [1623]). So suppose the c_i form a solution as described, and in addition that $c_{n+1} = 0$. Then we can write the *i*-th equation of system (**) as,

$$a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \dots + a_{in}c_n - b_i(0) = 0$$

which becomes

$$a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \dots + a_{in}c_n = 0$$

Since this is true for each *i*, we have that $x_1 = c_1, x_2 = c_2, x_3 = c_3, \ldots, x_n = c_n$ is a solution to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ formed with a nonsingular coefficient matrix. This means that the only possible solution is the trivial solution, so $c_1 = 0, c_2 = 0, c_3 = 0, \ldots, c_n = 0$. So, assuming simply that $c_{n+1} = 0$, we conclude that all of the c_i are zero. But this contradicts our choice of the c_i as not being the trivial solution to the system (**). So $c_{n+1} \neq 0$.

We now propose and verify a solution to the original system (*). Set

$$x_1 = \frac{c_1}{c_{n+1}}$$
 $x_2 = \frac{c_2}{c_{n+1}}$ $x_3 = \frac{c_3}{c_{n+1}}$... $x_n = \frac{c_n}{c_{n+1}}$

Notice how it was necessary that we know that $c_{n+1} \neq 0$ for this step to succeed. Now, evaluate the *i*-th equation of system (*) with this proposed solution, and recognize in the third line that c_1 through c_{n+1} appear as if they were substituted into the left-hand side of the *i*-th equation of system (**),

$$a_{i1}\frac{c_1}{c_{n+1}} + a_{i2}\frac{c_2}{c_{n+1}} + a_{i3}\frac{c_3}{c_{n+1}} + \dots + a_{in}\frac{c_n}{c_{n+1}}$$

= $\frac{1}{c_{n+1}}(a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \dots + a_{in}c_n)$
= $\frac{1}{c_{n+1}}(a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \dots + a_{in}c_n - b_ic_{n+1}) + b_i$
= $\frac{1}{c_{n+1}}(0) + b_i$

$$= b_i$$

Since this equation is true for every i, we have found a solution to system (*). To finish, we still need to establish that this solution is *unique*.

With one solution in hand, we will entertain the possibility of a second solution. So assume system (*) has two solutions,

$$x_1 = d_1$$
 $x_2 = d_2$ $x_3 = d_3$... $x_n = d_n$
 $x_1 = e_1$ $x_2 = e_2$ $x_3 = e_3$... $x_n = e_n$

Then,

$$(a_{i1}(d_1 - e_1) + a_{i2}(d_2 - e_2) + a_{i3}(d_3 - e_3) + \dots + a_{in}(d_n - e_n))$$

= $(a_{i1}d_1 + a_{i2}d_2 + a_{i3}d_3 + \dots + a_{in}d_n) - (a_{i1}e_1 + a_{i2}e_2 + a_{i3}e_3 + \dots + a_{in}e_n)$
= $b_i - b_i$
= 0

This is the *i*-th equation of the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ evaluated with $x_j = d_j - e_j$, $1 \leq j \leq n$. Since A is nonsingular, we must conclude that this solution is the trivial solution, and so $0 = d_j - e_j$, $1 \leq j \leq n$. That is, $d_j = e_j$ for all j and the two solutions are identical, meaning any solution to (*) is unique.

Notice that the proposed solution $(x_i = \frac{c_i}{c_{n+1}})$ appeared in this proof with no motivation whatsoever. This is just fine in a proof. A proof should *convince* you that a theorem is *true*.

It is your job to *read* the proof and be convinced of every assertion. Questions like "Where did that come from?" or "How would I think of that?" have no bearing on the *validity* of the proof.

Chapter V Vectors

We have worked extensively in the last chapter with matrices, and some with vectors. In this chapter we will develop the properties of vectors, while preparing to study vector spaces (Chapter VS [658]). Initially we will depart from our study of systems of linear equations, but in Section LC [221] we will forge a connection between linear combinations and systems of linear equations in Theorem SLSLC [229]. This connection will allow us to understand systems of linear equations at a higher level, while consequently discussing them less frequently.

Section VO Vector Operations

In this section we define some new operations involving vectors, and collect some basic properties of these operations. Begin by recalling our definition of a column vector as an ordered list of complex numbers, written vertically (Definition CV [52]). The collection of all possible vectors of a fixed size is a commonly used set, so we start with its definition.

Definition VSCV

Vector Space of Column Vectors

The vector space \mathbb{C}^m is the set of all column vectors (Definition CV [52]) of size m with entries from the set of complex numbers, \mathbb{C} .

(This definition contains Notation VSCV.)

When a set similar to this is defined using only column vectors where all the entries are from the real numbers, it is written as \mathbb{R}^m and is known as **Euclidean** *m*-space.

The term "vector" is used in a variety of different ways. We have defined it as an ordered list written vertically. It could simply be an ordered list of numbers, and written as (2, 3, -1, 6). Or it could be interpreted as a point in m dimensions, such as (3, 4, -2) representing a point in three dimensions relative to x, y and z axes. With an interpretation as a point, we can construct an arrow from the origin to the point which is consistent with the notion that a vector has direction and magnitude.

 \wedge

All of these ideas can be shown to be related and equivalent, so keep that in mind as you connect the ideas of this course with ideas from other disciplines. For now, we'll stick with the idea that a vector is just a list of numbers, in some particular order.

Subsection VEASM Vector Equality, Addition, Scalar Multiplication

We start our study of this set by first defining what it means for two vectors to be the same.

Definition CVE Column Vector Equality

Suppose that $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$. Then \mathbf{u} and \mathbf{v} are equal, written $\mathbf{u} = \mathbf{v}$ if

$$[\mathbf{u}]_i = [\mathbf{v}]_i \qquad \qquad 1 \le i \le m$$

(This definition contains Notation CVE.)

Now this may seem like a silly (or even stupid) thing to say so carefully. Of course two vectors are equal if they are equal for each corresponding entry! Well, this is not as silly as it appears. We will see a few occasions later where the obvious definition is *not* the right one. And besides, in doing mathematics we need to be very careful about making all the necessary definitions and making them unambiguous. And we've done that here.

 \triangle

Notice now that the symbol '=' is now doing triple-duty. We know from our earlier education what it means for two numbers (real or complex) to be equal, and we take this for granted. In Definition SE [1604] we defined what it meant for two sets to be equal. Now we have defined what it means for two vectors to be equal, and that definition builds on our definition for when two numbers are equal when we use the condition $u_i = v_i$ for all $1 \leq i \leq m$. So think carefully about your objects when you see an equal sign and think about just which notion of equality you have encountered. This will be especially important when you are asked to construct proofs whose conclusion states that two objects are equal.

OK, let's do an example of vector equality that begins to hint at the utility of this definition.

Example VESE Vector equality for a system of equations

Consider the system of linear equations in Archetype B [1652],

$$-7x_1 - 6x_2 - 12x_3 = -33$$

$$5x_1 + 5x_2 + 7x_3 = 24$$

$$x_1 + 4x_3 = 5$$

Note the use of three equals signs — each indicates an equality of numbers (the linear expressions are numbers when we evaluate them with fixed values of the variable quantities).

Now write the vector equality,

$$\begin{bmatrix} -7x_1 - 6x_2 - 12x_3\\ 5x_1 + 5x_2 + 7x_3\\ x_1 + 4x_3 \end{bmatrix} = \begin{bmatrix} -33\\ 24\\ 5 \end{bmatrix}$$

By Definition CVE [199], this *single* equality (of two column vectors) translates into *three* simultaneous equalities of numbers that form the system of equations. So with this new notion of vector equality we can become less reliant on referring to *systems* of *simultaneous* equations. There's more to vector equality than just this, but this is a good example for starters and we will develop it further. \boxtimes

We will now define two operations on the set \mathbb{C}^m . By this we mean well-defined procedures that somehow convert vectors into other vectors. Here are two of the most basic definitions of the entire course.

Definition CVA Column Vector Addition

Suppose that $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$. The sum of \mathbf{u} and \mathbf{v} is the vector $\mathbf{u} + \mathbf{v}$ defined by

$$[\mathbf{u} + \mathbf{v}]_i = [\mathbf{u}]_i + [\mathbf{v}]_i \qquad 1 \le i \le m$$

(This definition contains Notation CVA.)

So vector addition takes two vectors of the same size and combines them (in a natural way!) to create a new vector of the same size. Notice that this definition is required, even

 \wedge

if we agree that this is the obvious, right, natural or correct way to do it. Notice too that the symbol '+' is being recycled. We all know how to add *numbers*, but now we have the same symbol extended to double-duty and we use it to indicate how to add two new objects, vectors. And this definition of our new meaning is built on our previous meaning of addition via the expressions $u_i + v_i$. Think about your objects, especially when doing proofs. Vector addition is easy, here's an example from \mathbb{C}^4 .

Example VA Addition of two vectors in \mathbb{C}^4 If

$$\mathbf{u} = \begin{bmatrix} 2\\ -3\\ 4\\ 2 \end{bmatrix} \qquad \qquad \mathbf{v} = \begin{bmatrix} -1\\ 5\\ 2\\ -7 \end{bmatrix}$$

then

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} 2\\ -3\\ 4\\ 2 \end{bmatrix} + \begin{bmatrix} -1\\ 5\\ 2\\ -7 \end{bmatrix} = \begin{bmatrix} 2+(-1)\\ -3+5\\ 4+2\\ 2+(-7) \end{bmatrix} = \begin{bmatrix} 1\\ 2\\ 6\\ -5 \end{bmatrix}.$$

 \boxtimes

Our second operation takes two objects of different types, specifically a number and a vector, and combines them to create another vector. In this context we call a number a **scalar** in order to emphasize that it is not a vector.

Definition CVSM Column Vector Scalar Multiplication

Suppose $\mathbf{u} \in \mathbb{C}^m$ and $\alpha \in \mathbb{C}$, then the scalar multiple of \mathbf{u} by α is the vector $\alpha \mathbf{u}$ defined by

$$\left[\alpha \mathbf{u}\right]_{i} = \alpha \left[\mathbf{u}\right]_{i} \qquad 1 \le i \le m$$

(This definition contains Notation CVSM.)

Notice that we are doing a kind of multiplication here, but we are *defining* a new type, perhaps in what appears to be a natural way. We use juxtaposition (smashing two symbols together side-by-side) to denote this operation rather than using a symbol like we did with vector addition. So this can be another source of confusion. When two symbols are next to each other, are we doing regular old multiplication, the kind we've done for years, or are we doing scalar vector multiplication, the operation we just defined? Think about your objects — if the first object is a scalar, and the second is a vector, then it *must* be that we are doing our new operation, and the *result* of this operation will be another vector.

Notice how consistency in notation can be an aid here. If we write scalars as lower case Greek letters from the start of the alphabet (such as α, β, \ldots) and write vectors in bold Latin letters from the end of the alphabet ($\mathbf{u}, \mathbf{v}, \ldots$), then we have some hints about what type of objects we are working with. This can be a blessing *and* a curse, since when we go read another book about linear algebra, or read an application in another discipline (physics, economics, ...) the types of notation employed may be very different and hence unfamiliar.

 \triangle

Again, computationally, vector scalar multiplication is very easy.

Example CVSM Scalar multiplication in \mathbb{C}^5 If

$$\mathbf{u} = \begin{bmatrix} 3\\1\\-2\\4\\-1 \end{bmatrix}$$

and $\alpha = 6$, then

$$\alpha \mathbf{u} = 6 \begin{bmatrix} 3\\1\\-2\\4\\-1 \end{bmatrix} = \begin{bmatrix} 6(3)\\6(1)\\6(-2)\\6(4)\\6(-1) \end{bmatrix} = \begin{bmatrix} 18\\6\\-12\\24\\-6 \end{bmatrix}.$$

 \boxtimes

Vector addition and scalar multiplication are the most natural and basic operations to perform on vectors, so it should be easy to have your computational device form a linear combination. See: Computation VLC.MMA [1568] Computation VLC.TI86 [1579] Computation VLC.TI83 [1582] Computation VLC.SAGE [1589]

Subsection VSP Vector Space Properties

With definitions of vector addition and scalar multiplication we can state, and prove, several properties of each operation, and some properties that involve their interplay. We now collect ten of them here for later reference.

Theorem VSPCV

Vector Space Properties of Column Vectors

Suppose that \mathbb{C}^m is the set of column vectors of size m (Definition VSCV [198]) with addition and scalar multiplication as defined in Definition CVA [201] and Definition CVSM [203]. Then

- ACC Additive Closure, Column Vectors If $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$, then $\mathbf{u} + \mathbf{v} \in \mathbb{C}^m$.
- SCC Scalar Closure, Column Vectors If $\alpha \in \mathbb{C}$ and $\mathbf{u} \in \mathbb{C}^m$, then $\alpha \mathbf{u} \in \mathbb{C}^m$.
- CC Commutativity, Column Vectors If $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$, then $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
- AAC Additive Associativity, Column Vectors If $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^m$, then $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.

- ZC Zero Vector, Column Vectors There is a vector, 0, called the zero vector, such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for all $\mathbf{u} \in \mathbb{C}^m$.
- AIC Additive Inverses, Column Vectors If $\mathbf{u} \in \mathbb{C}^m$, then there exists a vector $-\mathbf{u} \in \mathbb{C}^m$ so that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
- SMAC Scalar Multiplication Associativity, Column Vectors If $\alpha, \beta \in \mathbb{C}$ and $\mathbf{u} \in \mathbb{C}^m$, then $\alpha(\beta \mathbf{u}) = (\alpha \beta) \mathbf{u}$.
- DVAC Distributivity across Vector Addition, Column Vectors If $\alpha \in \mathbb{C}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$, then $\alpha(\mathbf{u} + \mathbf{v}) = \alpha \mathbf{u} + \alpha \mathbf{v}$.
- DSAC Distributivity across Scalar Addition, Column Vectors If $\alpha, \beta \in \mathbb{C}$ and $\mathbf{u} \in \mathbb{C}^m$, then $(\alpha + \beta)\mathbf{u} = \alpha \mathbf{u} + \beta \mathbf{u}$.
- OC One, Column Vectors If $\mathbf{u} \in \mathbb{C}^m$, then $1\mathbf{u} = \mathbf{u}$.

Proof While some of these properties seem very obvious, they all require proof. However, the proofs are not very interesting, and border on tedious. We'll prove one version of distributivity very carefully, and you can test your proof-building skills on some of the others. We need to establish an equality, so we will do so by beginning with one side of the equality, apply various definitions and theorems (listed to the right of each step) to massage

 \square

the expression from the left into the expression on the right. Here we go with a proof of Property DSAC [206]. For $1 \le i \le m$,

$$\begin{split} \left[(\alpha + \beta) \mathbf{u} \right]_i &= (\alpha + \beta) \left[\mathbf{u} \right]_i & \text{Definition CVSM [203]} \\ &= \alpha \left[\mathbf{u} \right]_i + \beta \left[\mathbf{u} \right]_i & \text{Distributivity in } \mathbb{C} \\ &= \left[\alpha \mathbf{u} \right]_i + \left[\beta \mathbf{u} \right]_i & \text{Definition CVSM [203]} \\ &= \left[\alpha \mathbf{u} + \beta \mathbf{u} \right]_i & \text{Definition CVA [201]} \end{split}$$

Since the individual components of the vectors $(\alpha + \beta)\mathbf{u}$ and $\alpha \mathbf{u} + \beta \mathbf{u}$ are equal for all i, $1 \le i \le m$, Definition CVE [199] tells us the vectors are equal.

Many of the conclusions of our theorems can be characterized as "identities," especially when we are establishing basic properties of operations such as those in this section. Most of the properties listed in Theorem VSPCV [205] are examples. So some advice about the style we use for proving identities is appropriate right now. Have a look at Technique PI [1627].

Be careful with the notion of the vector $-\mathbf{u}$. This is a vector that we add to \mathbf{u} so that the result is the particular vector $\mathbf{0}$. This is basically a property of vector addition. It happens that we can compute $-\mathbf{u}$ using the *other* operation, scalar multiplication. We can prove this directly by writing that

$$[-\mathbf{u}]_i = - [\mathbf{u}]_i = (-1) [\mathbf{u}]_i = [(-1)\mathbf{u}]_i$$

We will see later how to derive this property as a *consequence* of several of the ten properties listed in Theorem VSPCV [205].

Similarly, we will often write something you would immediately recognize as "vector subtraction." This could be placed on a firm theoretical foundation — as you can do yourself with Exercise VO.T30 [213].

A final note. Property AAC [205] implies that we do not have to be careful about how we "parenthesize" the addition of vectors. In other words, there is nothing to be gained by writing $(\mathbf{u} + \mathbf{v}) + (\mathbf{w} + (\mathbf{x} + \mathbf{y}))$ rather than $\mathbf{u} + \mathbf{v} + \mathbf{w} + \mathbf{x} + \mathbf{y}$, since we get the same result no matter which order we choose to perform the four additions. So we won't be careful about using parentheses this way.

Subsection READ Reading Questions

- 1. Where have you seen vectors used before in other courses? How were they different?
- 2. In words, when are two vectors equal?

3. Perform the following computation with vector operations

$$2\begin{bmatrix}1\\5\\0\end{bmatrix} + (-3)\begin{bmatrix}7\\6\\5\end{bmatrix}$$

Subsection EXC Exercises

C10 Compute

$$4\begin{bmatrix}2\\-3\\4\\1\\0\end{bmatrix} + (-2)\begin{bmatrix}1\\2\\-5\\2\\4\end{bmatrix} + \begin{bmatrix}-1\\3\\0\\1\\2\end{bmatrix}$$

Contributed by Robert Beezer Solution [215]

C11 Solve the given vector equation for x, or explain why no solution exists:

$$3\begin{bmatrix}1\\2\\-1\end{bmatrix}+4\begin{bmatrix}2\\0\\x\end{bmatrix}=\begin{bmatrix}11\\6\\17\end{bmatrix}$$

Contributed by Chris Black Solution [215]

C12 Solve the given vector equation for α , or explain why no solution exists:

$$\alpha \begin{bmatrix} 1\\2\\-1 \end{bmatrix} + 4 \begin{bmatrix} 3\\4\\2 \end{bmatrix} = \begin{bmatrix} -1\\0\\4 \end{bmatrix}$$

Contributed by Chris Black Solution [215]

C13 Solve the given vector equation for α , or explain why no solution exists:

$$\alpha \begin{bmatrix} 3\\2\\-2 \end{bmatrix} + \begin{bmatrix} 6\\1\\2 \end{bmatrix} = \begin{bmatrix} 0\\-3\\6 \end{bmatrix}$$

Contributed by Chris Black Solution [216]

C14 Find α and β that solve the vector equation.

$$\alpha \begin{bmatrix} 1\\0 \end{bmatrix} + \beta \begin{bmatrix} 0\\1 \end{bmatrix} = \begin{bmatrix} 3\\2 \end{bmatrix}$$

Contributed by Chris Black Solution [217]

C15 Find α and β that solve the vector equation.

$$\alpha \begin{bmatrix} 2\\1 \end{bmatrix} + \beta \begin{bmatrix} 1\\3 \end{bmatrix} = \begin{bmatrix} 5\\0 \end{bmatrix}$$

Contributed by Chris Black Solution [217]

T05 Provide reasons (mostly vector space properties) as justification for each of the seven steps of the following proof.

Theorem For any vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^m$, if $\mathbf{u} + \mathbf{v} = \mathbf{u} + \mathbf{w}$, then $\mathbf{v} = \mathbf{w}$. **Proof:** Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^m$, and suppose $\mathbf{u} + \mathbf{v} = \mathbf{u} + \mathbf{w}$.

$$\mathbf{v} = \mathbf{0} + \mathbf{v}$$

= $(-\mathbf{u} + \mathbf{u}) + \mathbf{v}$
= $-\mathbf{u} + (\mathbf{u} + \mathbf{v})$
= $(-\mathbf{u} + (\mathbf{u} + \mathbf{w}))$
= $(-\mathbf{u} + \mathbf{u}) + \mathbf{w}$
= $\mathbf{0} + \mathbf{w}$
= \mathbf{w}

Contributed by Chris Black Solution [218]

T06 Provide reasons (mostly vector space properties) as justification for each of the six steps of the following proof.

Theorem For any vector $\mathbf{u} \in \mathbb{C}^m$, $0\mathbf{u} = \mathbf{0}$. **Proof:** Let $\mathbf{u} \in \mathbb{C}^m$.

$$0 = 0u + (-0u)$$

= (0 + 0)u + (-0u)
= (0u + 0u) + (-0u)
= 0u + (0u + (-0u))
= 0u + 0
= 0u

Contributed by Chris Black Solution [218]

T07 Provide reasons (mostly vector space properties) as justification for each of the six steps of the following proof.

Theorem For any scalar $c, c \mathbf{0} = \mathbf{0}$.

Proof: Let c be an arbitrary scalar.

$$0 = c0 + (-c0)$$

= c(0 + 0) + (-c0)
= (c0 + c0) + (-c0)
= c0 + (c0 + (-c0))
= c0 + 0

 $= c\mathbf{0}$

Contributed by Chris Black Solution [219]

T13 Prove Property CC [205] of Theorem VSPCV [205]. Write your proof in the style of the proof of Property DSAC [206] given in this section. Contributed by Robert Beezer Solution [219]

T17 Prove Property SMAC [206] of Theorem VSPCV [205]. Write your proof in the style of the proof of Property DSAC [206] given in this section. Contributed by Robert Beezer

T18 Prove Property DVAC [206] of Theorem VSPCV [205]. Write your proof in the style of the proof of Property DSAC [206] given in this section. Contributed by Robert Beezer

T30 Suppose **u** and **v** are two vectors in \mathbb{C}^m . Define a new operation, called "subtraction," as the new vector denoted $\mathbf{u} - \mathbf{v}$ and defined by

$$[\mathbf{u} - \mathbf{v}]_i = [\mathbf{u}]_i - [\mathbf{v}]_i \qquad 1 \le i \le m$$

Prove that we can express the subtraction of two vectors in terms of our two basic operations. More precisely, prove that $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$. So in a sense, subtraction is not something new and different, but is just a convenience. Mimic the style of similar proofs in this section. Contributed by Robert Beezer **T31** Review the definition of vector subtraction in Exercise VO.T30 [213]. Prove, by using counterexamples, that vector subtraction is not commutative and not associative. Contributed by Robert Beezer

T32 Review the definition of vector subtraction in Exercise VO.T30 [213]. Prove that vector subtraction obeys a distributive property. Specifically, prove that $\alpha (\mathbf{u} - \mathbf{v}) = \alpha \mathbf{u} - \alpha \mathbf{v}$.

Can you give two different proofs? Base one on the definition given in Exercise VO.T30 [213] and base the other on the equivalent formulation proved in Exercise VO.T30 [213]. Contributed by Robert Beezer

Subsection SOL Solutions

C10 Contributed by Robert Beezer Statement [209] $\begin{bmatrix} 5\\-13\\26\\1\\-6 \end{bmatrix}$

C11 Contributed by Chris Black Statement [209] Performing the indicated operations (Definition CVA [201], Definition CVSM [203]), we obtain the vector equations

$$\begin{bmatrix} 11\\6\\17 \end{bmatrix} = 3 \begin{bmatrix} 1\\2\\-1 \end{bmatrix} + 4 \begin{bmatrix} 2\\0\\x \end{bmatrix} = \begin{bmatrix} 11\\6\\-3+4x \end{bmatrix}$$

Since the entries of the vectors must be equal by Definition CVE [199], we have -3+4x = 17, which leads to x = 5.

C12 Contributed by Chris Black Statement [210] Performing the indicated operations (Definition CVA [201], Definition CVSM [203]), we obtain the vector equations

$$\begin{bmatrix} \alpha \\ 2\alpha \\ -\alpha \end{bmatrix} + \begin{bmatrix} 12 \\ 16 \\ 8 \end{bmatrix} = \begin{bmatrix} \alpha + 12 \\ 2\alpha + 16 \\ -\alpha + 8 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$$

Thus, if a solution α exists, by Definition CVE [199] then α must satisfy the three equations:

$$\alpha + 12 = -1$$
$$2\alpha + 16 = 0$$
$$-\alpha + 8 = 4$$

which leads to $\alpha = -13$, $\alpha = -8$ and $\alpha = 4$. Since α cannot simultaneously have three different values, there is no solution to the original vector equation.

C13 Contributed by Chris Black Statement [210] Performing the indicated operations (Definition CVA [201], Definition CVSM [203]), we obtain the vector equations

$$\begin{bmatrix} 3\alpha\\ 2\alpha\\ -2\alpha \end{bmatrix} + \begin{bmatrix} 6\\ 1\\ 2 \end{bmatrix} = \begin{bmatrix} 3\alpha+6\\ 2\alpha+1\\ -2\alpha+2 \end{bmatrix} = \begin{bmatrix} 0\\ -3\\ 6 \end{bmatrix}$$

Thus, if a solution α exists, by Definition CVE [199] then α must satisfy the three equations:

$$3\alpha + 6 = 0$$

 $2\alpha + 1 = -3$ $-2\alpha + 2 = 6$

which leads to $3\alpha = -6$, $2\alpha = -4$ and $-2\alpha = 4$. And thus, the solution to the given vector equation is $\alpha = -2$.

C14 Contributed by Chris Black Statement [210] Performing the indicated operations (Definition CVA [201], Definition CVSM [203]), we obtain the vector equations

$$\begin{bmatrix} 3\\ 2 \end{bmatrix} = \alpha \begin{bmatrix} 1\\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0\\ 1 \end{bmatrix} = \begin{bmatrix} \alpha + 0\\ 0 + \beta \end{bmatrix} = \begin{bmatrix} \alpha\\ \beta \end{bmatrix}$$

Since the entries of the vectors must be equal by Definition CVE [199], we have $\alpha = 3$ and $\beta = 2$.

C15 Contributed by Chris Black Statement [210] Performing the indicated operations (Definition CVA [201], Definition CVSM [203]), we obtain the vector equations

$$\begin{bmatrix} 5\\0 \end{bmatrix} = \alpha \begin{bmatrix} 2\\1 \end{bmatrix} + \beta \begin{bmatrix} 1\\3 \end{bmatrix} = \begin{bmatrix} 2\alpha + \beta\\\alpha + 3\beta \end{bmatrix}$$

Since the entries of the vectors must be equal by Definition CVE [199], we obtain the system of equations

$$2\alpha + \beta = 5$$

 $\alpha + 3\beta = 0.$

which we can solve by row-reducing the augmented matrix of the system,

2	1	5]	RREF	$\left[1 \right]$	0	3]
1	3	0	\longrightarrow	0	1	-1

Thus, the only solution is $\alpha = 3$, $\beta = -1$.

T05 Contributed by Chris Black Statement [211]

$$\mathbf{v} = \mathbf{0} + \mathbf{v}$$
Property ZC [206] $= (-\mathbf{u} + \mathbf{u}) + \mathbf{v}$ Property AIC [206] $= -\mathbf{u} + (\mathbf{u} + \mathbf{v})$ Property AAC [205] $= -\mathbf{u} + (\mathbf{u} + \mathbf{w})$ Hypothesis $= (-\mathbf{u} + \mathbf{u}) + \mathbf{w}$ Property AAC [205] $= \mathbf{0} + \mathbf{w}$ Property AIC [206] $= \mathbf{w}$ Property ZC [206]

T06 Contributed by Chris Black Statement [211]

 $\mathbf{0} = 0\mathbf{u} + (-0\mathbf{u}) \qquad \qquad \text{Property AIC [206]}$

[206]

$$= (0+0)\mathbf{u} + (-0\mathbf{u})$$
Property ZCN [1597] $= (0\mathbf{u} + 0\mathbf{u}) + (-0\mathbf{u})$ Property DVAC [206] $= 0\mathbf{u} + (0\mathbf{u} + (-0\mathbf{u}))$ Property AAC [205] $= 0\mathbf{u} + \mathbf{0}$ Property AIC [206] $= 0\mathbf{u}$ Property ZC [206]

T07Contributed by Chris Black Statement [212]

$$0 = c0 + (-c0)$$

= $c(0 + 0) + (-c0)$
= $(c0 + c0) + (-c0)$
= $c0 + (c0 + (-c0))$
= $c0 + 0$
= $c0$

Property AIC [206] Property ZC [206] Property DVAC [206] Property AAC [205] Property AIC [206] Property ZC [206]

T13 Statement [213] Contributed by Robert Beezer For all $1 \leq i \leq m$,

$$egin{aligned} \left[\mathbf{u}+\mathbf{v}
ight]_i &= \left[\mathbf{u}
ight]_i + \left[\mathbf{v}
ight]_i \ &= \left[\mathbf{v}
ight]_i + \left[\mathbf{u}
ight]_i \end{aligned}$$

Definition CVA [201] Commutativity in \mathbb{C}

$$= [\mathbf{v} + \mathbf{u}]_i$$
 Definition CVA [201]

With equality of each component of the vectors $\mathbf{u} + \mathbf{v}$ and $\mathbf{v} + \mathbf{u}$ being equal Definition CVE [199] tells us the two vectors are equal.

Section LC Linear Combinations

In Section VO [198] we defined vector addition and scalar multiplication. These two operations combine nicely to give us a construction known as a linear combination, a construct that we will work with throughout this course.

Subsection LC Linear Combinations

Definition LCCV Linear Combination of Column Vectors

Given *n* vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n$ from \mathbb{C}^m and *n* scalars $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$, their **linear** combination is the vector

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_n \mathbf{u}_n$$

 \triangle

So this definition takes an equal number of scalars and vectors, combines them using our

two new operations (scalar multiplication and vector addition) and creates a single brandnew vector, of the same size as the original vectors. When a definition or theorem employs a linear combination, think about the nature of the objects that go into its creation (lists of scalars and vectors), and the type of object that results (a single vector). Computationally, a linear combination is pretty easy.

Example TLC Two linear combinations in \mathbb{C}^6 Suppose that

and

 $\alpha_1 = 1 \qquad \alpha_2 = -4 \qquad \alpha_3 = 2 \qquad \alpha_4 = -1$ $\begin{bmatrix} 2 \\ -5 \end{bmatrix} \qquad \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$$\mathbf{u}_{1} = \begin{bmatrix} 2 \\ 4 \\ -3 \\ 1 \\ 2 \\ 9 \end{bmatrix} \qquad \mathbf{u}_{2} = \begin{bmatrix} 0 \\ 3 \\ 0 \\ -2 \\ 1 \\ 4 \end{bmatrix} \qquad \mathbf{u}_{3} = \begin{bmatrix} -3 \\ 2 \\ 1 \\ 1 \\ -3 \\ 0 \end{bmatrix} \qquad \mathbf{u}_{4} = \begin{bmatrix} 3 \\ 2 \\ -5 \\ 7 \\ 1 \\ 3 \end{bmatrix}$$

then their linear combination is

$$\alpha_{1}\mathbf{u}_{1} + \alpha_{2}\mathbf{u}_{2} + \alpha_{3}\mathbf{u}_{3} + \alpha_{4}\mathbf{u}_{4} = (1) \begin{bmatrix} 2\\4\\-3\\1\\2\\9 \end{bmatrix} + (-4) \begin{bmatrix} 6\\3\\0\\-2\\1\\4 \end{bmatrix} + (2) \begin{bmatrix} -5\\2\\1\\1\\-3\\0 \end{bmatrix} + (-1) \begin{bmatrix} 3\\2\\-5\\7\\1\\3 \end{bmatrix}$$
$$= \begin{bmatrix} 2\\4\\-3\\1\\2\\9 \end{bmatrix} + \begin{bmatrix} -24\\-12\\0\\8\\-4\\-16 \end{bmatrix} + \begin{bmatrix} -10\\4\\2\\2\\-6\\0 \end{bmatrix} + \begin{bmatrix} -3\\-2\\5\\-7\\-1\\-3 \end{bmatrix} = \begin{bmatrix} -35\\-6\\4\\4\\-9\\-10 \end{bmatrix}$$

A different linear combination, of the same set of vectors, can be formed with different scalars. Take

$$\beta_1 = 3 \qquad \qquad \beta_2 = 0 \qquad \qquad \beta_3 = 5 \qquad \qquad \beta_4 = -1$$

and form the linear combination

$$\beta_{1}\mathbf{u}_{1} + \beta_{2}\mathbf{u}_{2} + \beta_{3}\mathbf{u}_{3} + \beta_{4}\mathbf{u}_{4} = (3)\begin{bmatrix}2\\4\\-3\\1\\2\\9\end{bmatrix} + (0)\begin{bmatrix}6\\3\\0\\-2\\1\\4\end{bmatrix} + (5)\begin{bmatrix}-5\\2\\1\\1\\-3\\0\end{bmatrix} + (-1)\begin{bmatrix}3\\2\\-5\\7\\1\\3\end{bmatrix}$$
$$= \begin{bmatrix}6\\12\\-9\\3\\6\\27\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} + \begin{bmatrix}-25\\10\\5\\5\\-7\\-1\\5\\0\end{bmatrix} + \begin{bmatrix}-3\\-2\\5\\-7\\-7\\-1\\-3\end{bmatrix} = \begin{bmatrix}-22\\20\\1\\1\\-10\\24\end{bmatrix}$$

Notice how we could keep our set of vectors fixed, and use different sets of scalars to construct different vectors. You might build a few new linear combinations of \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 , \mathbf{u}_4 right now. We'll be right here when you get back. What vectors were you able to create? Do you think

you could create the vector

$$\mathbf{w} = \begin{bmatrix} 13\\15\\5\\-17\\2\\25 \end{bmatrix}$$

with a "suitable" choice of four scalars? Do you think you could create *any* possible vector from \mathbb{C}^6 by choosing the proper scalars? These last two questions are very fundamental, and time spent considering them *now* will prove beneficial later.

Our next two examples are key ones, and a discussion about decompositions is timely. Have a look at Technique DC [1628] before studying the next two examples.

Example ABLC Archetype B as a linear combination

In this example we will rewrite Archetype B [1652] in the language of vectors, vector equality and linear combinations. In Example VESE [200] we wrote the system of m = 3 equations as the vector equality

$$\begin{bmatrix} -7x_1 - 6x_2 - 12x_3\\ 5x_1 + 5x_2 + 7x_3\\ x_1 + 4x_3 \end{bmatrix} = \begin{bmatrix} -33\\ 24\\ 5 \end{bmatrix}$$

Now we will bust up the linear expressions on the left, first using vector addition,

$$\begin{bmatrix} -7x_1\\5x_1\\x_1 \end{bmatrix} + \begin{bmatrix} -6x_2\\5x_2\\0x_2 \end{bmatrix} + \begin{bmatrix} -12x_3\\7x_3\\4x_3 \end{bmatrix} = \begin{bmatrix} -33\\24\\5 \end{bmatrix}$$

Now we can rewrite each of these n = 3 vectors as a scalar multiple of a fixed vector, where the scalar is one of the unknown variables, converting the left-hand side into a linear combination

$$x_1 \begin{bmatrix} -7\\5\\1 \end{bmatrix} + x_2 \begin{bmatrix} -6\\5\\0 \end{bmatrix} + x_3 \begin{bmatrix} -12\\7\\4 \end{bmatrix} = \begin{bmatrix} -33\\24\\5 \end{bmatrix}$$

We can now interpret the problem of solving the system of equations as determining values for the scalar multiples that make the vector equation true. In the analysis of Archetype B [1652], we were able to determine that it had only one solution. A quick way to see this is to row-reduce the coefficient matrix to the 3×3 identity matrix and apply Theorem NMRRI [176] to determine that the coefficient matrix is nonsingular. Then Theorem NMUS [180] tells us that the system of equations has a unique solution. This solution is

$$x_1 = -3$$
 $x_2 = 5$ $x_3 = 2$

So, in the context of this example, we can express the fact that these values of the variables

are a solution by writing the linear combination,

$$(-3)\begin{bmatrix}-7\\5\\1\end{bmatrix} + (5)\begin{bmatrix}-6\\5\\0\end{bmatrix} + (2)\begin{bmatrix}-12\\7\\4\end{bmatrix} = \begin{bmatrix}-33\\24\\5\end{bmatrix}$$

Furthermore, these are the only three scalars that will accomplish this equality, since they come from a unique solution.

Notice how the three vectors in this example are the columns of the coefficient matrix of the system of equations. This is our first hint of the important interplay between the vectors that form the columns of a matrix, and the matrix itself. \boxtimes

With any discussion of Archetype A [1642] or Archetype B [1652] we should be sure to contrast with the other.

Example AALC Archetype A as a linear combination

As a vector equality, Archetype A [1642] can be written as

$$\begin{bmatrix} x_1 - x_2 + 2x_3 \\ 2x_1 + x_2 + x_3 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 8 \\ 5 \end{bmatrix}$$

Now bust up the linear expressions on the left, first using vector addition,

$$\begin{bmatrix} x_1\\2x_1\\x_1\end{bmatrix} + \begin{bmatrix} -x_2\\x_2\\x_2\end{bmatrix} + \begin{bmatrix} 2x_3\\x_3\\0x_3\end{bmatrix} = \begin{bmatrix} 1\\8\\5\end{bmatrix}$$

Rewrite each of these n = 3 vectors as a scalar multiple of a fixed vector, where the scalar is one of the unknown variables, converting the left-hand side into a linear combination

$$x_1 \begin{bmatrix} 1\\2\\1 \end{bmatrix} + x_2 \begin{bmatrix} -1\\1\\1 \end{bmatrix} + x_3 \begin{bmatrix} 2\\1\\0 \end{bmatrix} = \begin{bmatrix} 1\\8\\5 \end{bmatrix}$$

Row-reducing the augmented matrix for Archetype A [1642] leads to the conclusion that the system is consistent and has free variables, hence infinitely many solutions. So for example, the two solutions

$$x_1 = 2$$
 $x_2 = 3$ $x_3 = 1$
 $x_1 = 3$ $x_2 = 2$ $x_3 = 0$

can be used together to say that,

$$(2)\begin{bmatrix}1\\2\\1\end{bmatrix} + (3)\begin{bmatrix}-1\\1\\1\end{bmatrix} + (1)\begin{bmatrix}2\\1\\0\end{bmatrix} = \begin{bmatrix}1\\8\\5\end{bmatrix} = (3)\begin{bmatrix}1\\2\\1\end{bmatrix} + (2)\begin{bmatrix}-1\\1\\1\end{bmatrix} + (0)\begin{bmatrix}2\\1\\0\end{bmatrix}$$

Ignore the middle of this equation, and move all the terms to the left-hand side,

$$(2)\begin{bmatrix}1\\2\\1\end{bmatrix} + (3)\begin{bmatrix}-1\\1\\1\end{bmatrix} + (1)\begin{bmatrix}2\\1\\0\end{bmatrix} + (-3)\begin{bmatrix}1\\2\\1\end{bmatrix} + (-2)\begin{bmatrix}-1\\1\\1\end{bmatrix} + (-0)\begin{bmatrix}2\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

Regrouping gives

$$(-1)\begin{bmatrix}1\\2\\1\end{bmatrix} + (1)\begin{bmatrix}-1\\1\\1\end{bmatrix} + (1)\begin{bmatrix}2\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

Notice that these three vectors are the columns of the coefficient matrix for the system of equations in Archetype A [1642]. This equality says there is a linear combination of those columns that equals the vector of all zeros. Give it some thought, but this says that

$$x_1 = -1$$
 $x_2 = 1$ $x_3 = 1$

is a nontrivial solution to the homogeneous system of equations with the coefficient matrix for the original system in Archetype A [1642]. In particular, this demonstrates that this coefficient matrix is singular. \boxtimes

There's a lot going on in the last two examples. Come back to them in a while and make some connections with the intervening material. For now, we will summarize and explain some of this behavior with a theorem.

Theorem SLSLC

Solutions to Linear Systems are Linear Combinations

Denote the columns of the $m \times n$ matrix A as the vectors $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n$. Then $\mathbf{x} \in \mathbb{C}n$ is a solution to the linear system of equations $\mathcal{LS}(A, \mathbf{b})$ if and only if \mathbf{b} equals the linear combination of the columns of A formed with the entries of \mathbf{x} ,

$$[\mathbf{x}]_1 \mathbf{A}_1 + [\mathbf{x}]_2 \mathbf{A}_2 + [\mathbf{x}]_3 \mathbf{A}_3 + \dots + [\mathbf{x}]_n \mathbf{A}_n = \mathbf{b}$$

Proof The proof of this theorem is as much about a change in notation as it is about making logical deductions. Write the system of equations $\mathcal{LS}(A, \mathbf{b})$ as

 $a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n = b_1$ $a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n = b_2$ $a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3n}x_n = b_3$ \vdots $a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n = b_m$

Notice then that the entry of the coefficient matrix A in row i and column j has two names: a_{ij} as the coefficient of x_j in equation i of the system and $[\mathbf{A}_j]_i$ as the *i*-th entry of the column vector in column j of the coefficient matrix A. Likewise, entry i of \mathbf{b} has two names: b_i from the linear system and $[\mathbf{b}]_i$ as an entry of a vector. Our theorem is an equivalence (Technique E [1619]) so we need to prove both "directions."

(\Leftarrow) Suppose we have the vector equality between **b** and the linear combination of the columns of A. Then for $1 \le i \le m$,

$$b_{i} = [\mathbf{b}]_{i} \qquad \text{Notation CVC [52]}$$

$$= [[\mathbf{x}]_{1} \mathbf{A}_{1} + [\mathbf{x}]_{2} \mathbf{A}_{2} + [\mathbf{x}]_{3} \mathbf{A}_{3} + \dots + [\mathbf{x}]_{n} \mathbf{A}_{n}]_{i} \qquad \text{Hypothesis}$$

$$= [[\mathbf{x}]_{1} \mathbf{A}_{1}]_{i} + [[\mathbf{x}]_{2} \mathbf{A}_{2}]_{i} + [[\mathbf{x}]_{3} \mathbf{A}_{3}]_{i} + \dots + [[\mathbf{x}]_{n} \mathbf{A}_{n}]_{i} \qquad \text{Definition CVA [201]}$$

$$= [\mathbf{x}]_{1} [\mathbf{A}_{1}]_{i} + [\mathbf{x}]_{2} [\mathbf{A}_{2}]_{i} + [\mathbf{x}]_{3} [\mathbf{A}_{3}]_{i} + \dots + [\mathbf{x}]_{n} [\mathbf{A}_{n}]_{i} \qquad \text{Definition CVSM [203]}$$

$$= [\mathbf{x}]_{1} a_{i1} + [\mathbf{x}]_{2} a_{i2} + [\mathbf{x}]_{3} a_{i3} + \dots + [\mathbf{x}]_{n} a_{in}$$
Notation CVC [52]
$$= a_{i1} [\mathbf{x}]_{1} + a_{i2} [\mathbf{x}]_{2} + a_{i3} [\mathbf{x}]_{3} + \dots + a_{in} [\mathbf{x}]_{n}$$
Property CMCN [1597]

This says that the entries of \mathbf{x} form a solution to equation i of $\mathcal{LS}(A, \mathbf{b})$ for all $1 \le i \le m$, in other words, \mathbf{x} is a solution to $\mathcal{LS}(A, \mathbf{b})$.

(⇒) Suppose now that **x** is a solution to the linear system $\mathcal{LS}(A, \mathbf{b})$. Then for all $1 \leq i \leq m$,

$$\begin{aligned} [\mathbf{b}]_{i} &= b_{i} & \text{Notation CVC [52]} \\ &= a_{i1} [\mathbf{x}]_{1} + a_{i2} [\mathbf{x}]_{2} + a_{i3} [\mathbf{x}]_{3} + \dots + a_{in} [\mathbf{x}]_{n} & \text{Hypothesis} \\ &= [\mathbf{x}]_{1} a_{i1} + [\mathbf{x}]_{2} a_{i2} + [\mathbf{x}]_{3} a_{i3} + \dots + [\mathbf{x}]_{n} a_{in} & \text{Property CMCN [1597]} \\ &= [\mathbf{x}]_{1} [\mathbf{A}_{1}]_{i} + [\mathbf{x}]_{2} [\mathbf{A}_{2}]_{i} + [\mathbf{x}]_{3} [\mathbf{A}_{3}]_{i} + \dots + [\mathbf{x}]_{n} [\mathbf{A}_{n}]_{i} & \text{Notation CVC [52]} \\ &= [[\mathbf{x}]_{1} \mathbf{A}_{1}]_{i} + [[\mathbf{x}]_{2} \mathbf{A}_{2}]_{i} + [[\mathbf{x}]_{3} \mathbf{A}_{3}]_{i} + \dots + [[\mathbf{x}]_{n} \mathbf{A}_{n}]_{i} & \text{Definition CVSM [203]} \\ &= [[\mathbf{x}]_{1} \mathbf{A}_{1} + [\mathbf{x}]_{2} \mathbf{A}_{2} + [\mathbf{x}]_{3} \mathbf{A}_{3} + \dots + [\mathbf{x}]_{n} \mathbf{A}_{n}]_{i} & \text{Definition CVA [201]} \end{aligned}$$

Since the components of **b** and the linear combination of the columns of A agree for all $1 \le i \le m$, Definition CVE [199] tells us that the vectors are equal.

In other words, this theorem tells us that solutions to systems of equations are linear combinations of the *n* column vectors of the coefficient matrix (\mathbf{A}_j) which yield the constant vector **b**. Or said another way, a solution to a system of equations $\mathcal{LS}(A, \mathbf{b})$ is an answer to the question "How can I form the vector **b** as a linear combination of the columns of A?" Look through the archetypes that are systems of equations and examine a few of the

advertised solutions. In each case use the solution to form a linear combination of the columns of the coefficient matrix and verify that the result equals the constant vector (see Exercise LC.C21 [263]).

Subsection VFSS Vector Form of Solution Sets

We have written solutions to systems of equations as column vectors. For example Archetype B [1652] has the solution $x_1 = -3$, $x_2 = 5$, $x_3 = 2$ which we now write as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 5 \\ 2 \end{bmatrix}$$

Now, we will use column vectors and linear combinations to express *all* of the solutions to a linear system of equations in a compact and understandable way. First, here's two examples that will motivate our next theorem. This is a valuable technique, almost the equal of row-reducing a matrix, so be sure you get comfortable with it over the course of this section.

Example VFSAD Vector form of solutions for Archetype D

Archetype D [1670] is a linear system of 3 equations in 4 variables. Row-reducing the

augmented matrix yields

$$\begin{bmatrix} 1 & 0 & 3 & -2 & 4 \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and we see r = 2 nonzero rows. Also, $D = \{1, 2\}$ so the dependent variables are then x_1 and x_2 . $F = \{3, 4, 5\}$ so the two free variables are x_3 and x_4 . We will express a generic solution for the system by two slightly different methods, though both arrive at the same conclusion.

First, we will decompose (Technique DC [1628]) a solution vector. Rearranging each equation represented in the row-reduced form of the augmented matrix by solving for the dependent variable in each row yields the vector equality,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 - 3x_3 + 2x_4 \\ -x_3 + 3x_4 \\ x_3 \\ x_4 \end{bmatrix}$$

Now we will use the definitions of column vector addition and scalar multiplication to express this vector as a linear combination,

$$= \begin{bmatrix} 4\\0\\0\\0\end{bmatrix} + \begin{bmatrix} -3x_3\\-x_3\\x_3\\0\end{bmatrix} + \begin{bmatrix} 2x_4\\3x_4\\0\\x_4\end{bmatrix}$$

Definition CVA [201]

$$= \begin{bmatrix} 4\\0\\0\\0 \end{bmatrix} + x_3 \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix} + x_4 \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix}$$

Definition CVSM [203]

We will develop the same linear combination a bit quicker, using three steps. While the method above is instructive, the method below will be our preferred approach.

Step 1. Write the vector of variables as a fixed vector, plus a linear combination of n-r vectors, using the free variables as the scalars.

Step 2. Use 0's and 1's to ensure equality for the entries of the vectors with indices in F (corresponding to the free variables).

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Step 3. For each dependent variable, use the augmented matrix to formulate an equation expressing the dependent variable as a constant plus multiples of the free variables. Convert

this equation into entries of the vectors that ensure equality for each dependent variable, one at a time.

$$x_{1} = 4 - 3x_{3} + 2x_{4} \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$
$$x_{2} = 0 - 1x_{3} + 3x_{4} \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

This final *form* of a typical solution is especially pleasing and useful. For example, we can build solutions quickly by choosing values for our free variables, and then compute a linear combination. Such as

$$x_{3} = 2, x_{4} = -5 \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (2) \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + (-5) \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -12 \\ -17 \\ 2 \\ -5 \end{bmatrix}$$

or,

$$x_{3} = 1, x_{4} = 3 \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (1) \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + (3) \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 1 \\ 3 \end{bmatrix}$$

You'll find the second solution listed in the write-up for Archetype D [1670], and you might check the first solution by substituting it back into the original equations.

While this form is useful for quickly creating solutions, it's even better because it tells us *exactly* what every solution looks like. We know the solution set is infinite, which is pretty

big, but now we can say that a solution is some multiple of $\begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix}$ plus a multiple of $\begin{bmatrix} 2\\3\\0\\1 \end{bmatrix}$

plus the fixed vector $\begin{bmatrix} 4\\0\\0\\0 \end{bmatrix}$. Period. So it only takes us *three* vectors to describe the entire infinite solution set, provided we also agree on how to combine the three vectors into a linear combination.

This is such an important and fundamental technique, we'll do another example.

Example VFS

Vector form of solutions

Consider a linear system of m = 5 equations in n = 7 variables, having the augmented matrix A.

$$A = \begin{bmatrix} 2 & 1 & -1 & -2 & 2 & 1 & 5 & 21 \\ 1 & 1 & -3 & 1 & 1 & 1 & 2 & -5 \\ 1 & 2 & -8 & 5 & 1 & 1 & -6 & -15 \\ 3 & 3 & -9 & 3 & 6 & 5 & 2 & -24 \\ -2 & -1 & 1 & 2 & 1 & 1 & -9 & -30 \end{bmatrix}$$

Row-reducing we obtain the matrix

$$B = \begin{bmatrix} 1 & 0 & 2 & -3 & 0 & 0 & 9 & 15 \\ 0 & 1 & -5 & 4 & 0 & 0 & -8 & -10 \\ 0 & 0 & 0 & 0 & 1 & 0 & -6 & 11 \\ 0 & 0 & 0 & 0 & 0 & 1 & 7 & -21 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and we see r = 4 nonzero rows. Also, $D = \{1, 2, 5, 6\}$ so the dependent variables are then x_1, x_2, x_5 , and x_6 . $F = \{3, 4, 7, 8\}$ so the n - r = 3 free variables are x_3, x_4 and x_7 . We will express a generic solution for the system by two different methods: both a decomposition and a construction.

First, we will decompose (Technique DC [1628]) a solution vector. Rearranging each equation represented in the row-reduced form of the augmented matrix by solving for the

dependent variable in each row yields the vector equality,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 15 - 2x_3 + 3x_4 - 9x_7 \\ -10 + 5x_3 - 4x_4 + 8x_7 \\ x_3 \\ x_4 \\ 11 + 6x_7 \\ -21 - 7x_7 \\ x_7 \end{bmatrix}$$

Now we will use the definitions of column vector addition and scalar multiplication to decompose this generic solution vector as a linear combination,

$$= \begin{bmatrix} 15\\ -10\\ 0\\ 0\\ 11\\ -21\\ 0 \end{bmatrix} + \begin{bmatrix} -2x_3\\ 5x_3\\ x_3\\ 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} + \begin{bmatrix} 3x_4\\ -4x_4\\ 0\\ x_4\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} + \begin{bmatrix} -9x_7\\ 8x_7\\ 0\\ 0\\ 0\\ 6x_7\\ -7x_7\\ x_7 \end{bmatrix}$$

Definition CVA [201]

$$= \begin{bmatrix} 15\\ -10\\ 0\\ 0\\ 11\\ -21\\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -2\\ 5\\ 1\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3\\ -4\\ 0\\ 1\\ 0\\ 0\\ 0 \end{bmatrix} + x_7 \begin{bmatrix} -9\\ 8\\ 0\\ 0\\ 0\\ 6\\ -7\\ 1 \end{bmatrix}$$
 Definition CVSM [203]

We will now develop the same linear combination a bit quicker, using three steps. While the method above is instructive, the method below will be our preferred approach.

Step 1. Write the vector of variables as a fixed vector, plus a linear combination of n-r vectors, using the free variables as the scalars.

Step 2. Use 0's and 1's to ensure equality for the entries of the vectors with indices in F

(corresponding to the free variables).

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + x_7 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Step 3. For each dependent variable, use the augmented matrix to formulate an equation expressing the dependent variable as a constant plus multiples of the free variables. Convert this equation into entries of the vectors that ensure equality for each dependent variable, one at a time.

$$x_{1} = 15 - 2x_{3} + 3x_{4} - 9x_{7} \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 15 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\begin{aligned} x_{2} &= -10 + 5x_{3} - 4x_{4} + 8x_{7} \implies \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \\ x_{5} &= \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 11 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \\ x_{6} &= -21 - 7x_{7} \implies \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 11 \\ -21 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{3} \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -2 \\ -2 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -2 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4 \\ 0 \\ 0 \end{bmatrix} \\ x_{7} &= \begin{bmatrix} -9 \\ -4$$

This final *form* of a typical solution is especially pleasing and useful. For example, we can

build solutions quickly by choosing values for our free variables, and then compute a linear combination. For example

or perhaps,

$$\mathbf{x}_{3} = 5, \ x_{4} = 2, \ x_{7} = 1 \qquad \Rightarrow$$

$$\mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 0 \\ 11 \\ -21 \\ 0 \end{bmatrix} + (5) \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (2) \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + (1) \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 0 \\ 6 \\ -7 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 15 \\ 5 \\ 2 \\ 17 \\ -28 \\ 1 \end{bmatrix}$$

or even,

$$x_3 = 0, x_4 = 0, x_7 = 0 \qquad \Rightarrow \qquad$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 0 \\ 11 \\ -21 \\ 0 \end{bmatrix} + (0) \begin{bmatrix} -2 \\ 5 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (0) \begin{bmatrix} 3 \\ -4 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (0) \begin{bmatrix} -9 \\ 8 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 15 \\ -10 \\ 0 \\ 0 \\ 0 \\ 11 \\ -21 \\ 0 \end{bmatrix}$$

So we can compactly express *all* of the solutions to this linear system with just 4 fixed vectors, provided we agree how to combine them in a linear combinations to create solution vectors.

Suppose you were told that the vector \mathbf{w} below was a solution to this system of equations. Could you turn the problem around and write \mathbf{w} as a linear combination of the four vectors \mathbf{c} , \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 ? (See Exercise LC.M11 [266].)

$$\mathbf{w} = \begin{bmatrix} 100\\ -75\\ 7\\ 9\\ -37\\ 35\\ -8 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} 15\\ -10\\ 0\\ 0\\ 0\\ 11\\ -21\\ 0 \end{bmatrix} \qquad \mathbf{u}_1 = \begin{bmatrix} -2\\ 5\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} 3\\ -4\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} \qquad \mathbf{u}_3 = \begin{bmatrix} -9\\ 8\\ 0\\ 0\\ 0\\ 6\\ -7\\ 1 \end{bmatrix}$$

Did you think a few weeks ago that you could so quickly and easily list *all* the solutions

 \boxtimes

to a linear system of 5 equations in 7 variables?

We'll now formalize the last two (important) examples as a theorem.

Theorem VFSLS

Vector Form of Solutions to Linear Systems

Suppose that $[A \mid \mathbf{b}]$ is the augmented matrix for a consistent linear system $\mathcal{LS}(A, \mathbf{b})$ of m equations in n variables. Let B be a row-equivalent $m \times (n + 1)$ matrix in reduced rowechelon form. Suppose that B has r nonzero rows, columns without leading 1's with indices $F = \{f_1, f_2, f_3, \ldots, f_{n-r}, n+1\}$, and columns with leading 1's (pivot columns) having indices $D = \{d_1, d_2, d_3, \ldots, d_r\}$. Define vectors $\mathbf{c}, \mathbf{u}_j, 1 \le j \le n-r$ of size n by

$$[\mathbf{c}]_{i} = \begin{cases} 0 & \text{if } i \in F \\ [B]_{k,n+1} & \text{if } i \in D, \ i = d_{k} \end{cases}$$
$$[\mathbf{u}_{j}]_{i} = \begin{cases} 1 & \text{if } i \in F, \ i = f_{j} \\ 0 & \text{if } i \in F, \ i \neq f_{j} \\ -[B]_{k,f_{j}} & \text{if } i \in D, \ i = d_{k} \end{cases}$$

Then the set of solutions to the system of equations $\mathcal{LS}(A, \mathbf{b})$ is

$$S = \{ \mathbf{c} + \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_{n-r} \mathbf{u}_{n-r} | \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{n-r} \in \mathbb{C} \}$$

Proof First, $\mathcal{LS}(A, \mathbf{b})$ is equivalent to the linear system of equations that has the matrix B as its augmented matrix (Theorem REMES [61]), so we need only show that S is the

 \square

solution set for the system with B as its augmented matrix. The conclusion of this theorem is that the solution set is equal to the set S, so we will apply Definition SE [1604].

We begin by showing that every element of S is indeed a solution to the system. Let $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{n-r}$ be one choice of the scalars used to describe elements of S. So an arbitrary element of S, which we will consider as a proposed solution is

$$\mathbf{x} = \mathbf{c} + \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_{n-r} \mathbf{u}_{n-r}$$

When $r + 1 \leq \ell \leq m$, row ℓ of the matrix B is a zero row, so the equation represented by that row is always true, no matter which solution vector we propose. So concentrate on rows representing equations $1 \leq \ell \leq r$. We evaluate equation ℓ of the system represented by B with the proposed solution vector \mathbf{x} and refer to the value of the left-hand side of the equation as β_{ℓ} ,

$$\beta_{\ell} = [B]_{\ell 1} [\mathbf{x}]_1 + [B]_{\ell 2} [\mathbf{x}]_2 + [B]_{\ell 3} [\mathbf{x}]_3 + \dots + [B]_{\ell n} [\mathbf{x}]_n$$

Since $[B]_{\ell d_i} = 0$ for all $1 \le i \le r$, except that $[B]_{\ell d_\ell} = 1$, we see that β_ℓ simplifies to

$$\beta_{\ell} = [\mathbf{x}]_{d_{\ell}} + [B]_{\ell f_1} [\mathbf{x}]_{f_1} + [B]_{\ell f_2} [\mathbf{x}]_{f_2} + [B]_{\ell f_3} [\mathbf{x}]_{f_3} + \dots + [B]_{\ell f_{n-r}} [\mathbf{x}]_{f_{n-r}}$$

Notice that for $1 \leq i \leq n - r$

$$[\mathbf{x}]_{f_i} = [\mathbf{c}]_{f_i} + \alpha_1 [\mathbf{u}_1]_{f_i} + \alpha_2 [\mathbf{u}_2]_{f_i} + \alpha_3 [\mathbf{u}_3]_{f_i} + \dots + \alpha_i [\mathbf{u}_i]_{f_i} + \dots + \alpha_{n-r} [\mathbf{u}_{\mathbf{n}-\mathbf{r}}]_{f_i} = 0 + \alpha_1(0) + \alpha_2(0) + \alpha_3(0) + \dots + \alpha_i(1) + \dots + \alpha_{n-r}(0)$$

 $= \alpha_i$

So β_{ℓ} simplifies further, and we expand the first term

$$\begin{aligned} \beta_{\ell} &= [\mathbf{x}]_{d_{\ell}} + [B]_{\ell f_{1}} \alpha_{1} + [B]_{\ell f_{2}} \alpha_{2} + [B]_{\ell f_{3}} \alpha_{3} + \dots + [B]_{\ell f_{n-r}} \alpha_{n-r} \\ &= [\mathbf{c} + \alpha_{1} \mathbf{u}_{1} + \alpha_{2} \mathbf{u}_{2} + \alpha_{3} \mathbf{u}_{3} + \dots + \alpha_{n-r} \mathbf{u}_{n-r}]_{d_{\ell}} + \\ & [B]_{\ell f_{1}} \alpha_{1} + [B]_{\ell f_{2}} \alpha_{2} + [B]_{\ell f_{3}} \alpha_{3} + \dots + [B]_{\ell f_{n-r}} \alpha_{n-r} \\ &= [\mathbf{c}]_{d_{\ell}} + \alpha_{1} [\mathbf{u}_{1}]_{d_{\ell}} + \alpha_{2} [\mathbf{u}_{2}]_{d_{\ell}} + \alpha_{3} [\mathbf{u}_{3}]_{d_{\ell}} + \dots + \alpha_{n-r} [\mathbf{u}_{\mathbf{n-r}}]_{d_{\ell}} + \\ & [B]_{\ell f_{1}} \alpha_{1} + [B]_{\ell f_{2}} \alpha_{2} + [B]_{\ell f_{3}} \alpha_{3} + \dots + [B]_{\ell f_{n-r}} \alpha_{n-r} \\ &= [B]_{\ell,n+1} + \alpha_{1} (-[B]_{\ell,f_{1}}) + \alpha_{2} (-[B]_{\ell,f_{2}}) + \alpha_{3} (-[B]_{\ell,f_{3}}) + \dots + \alpha_{n-r} (-[B]_{\ell,f_{n-r}}) + \\ & [B]_{\ell f_{1}} \alpha_{1} + [B]_{\ell f_{2}} \alpha_{2} + [B]_{\ell f_{3}} \alpha_{3} + \dots + [B]_{\ell f_{n-r}} \alpha_{n-r} \\ &= [B]_{\ell,n+1} \end{aligned}$$

So β_{ℓ} began as the left-hand side of equation ℓ of the system represented by B and we now know it equals $[B]_{\ell,n+1}$, the constant term for equation ℓ of this system. So the arbitrarily chosen vector from S makes every equation of the system true, and therefore is a solution to the system. So all the elements of S are solutions to the system.

For the second half of the proof, assume that \mathbf{x} is a solution vector for the system having B as its augmented matrix. For convenience and clarity, denote the entries of \mathbf{x} by x_i , in other words, $x_i = [\mathbf{x}]_i$. We desire to show that this solution vector is also an element of the set S. Begin with the observation that a solution vector's entries makes equation ℓ of the

system true for all $1 \leq \ell \leq m$,

$$[B]_{\ell,1} x_1 + [B]_{\ell,2} x_2 + [B]_{\ell,3} x_3 + \dots + [B]_{\ell,n} x_n = [B]_{\ell,n+1}$$

When $\ell \leq r$, the pivot columns of *B* have zero entries in row ℓ with the exception of column d_{ℓ} , which will contain a 1. So for $1 \leq \ell \leq r$, equation ℓ simplifies to

$$1x_{d_{\ell}} + [B]_{\ell,f_1} x_{f_1} + [B]_{\ell,f_2} x_{f_2} + [B]_{\ell,f_3} x_{f_3} + \dots + [B]_{\ell,f_{n-r}} x_{f_{n-r}} = [B]_{\ell,n+1}$$

This allows us to write,

$$\begin{aligned} [\mathbf{x}]_{d_{\ell}} &= x_{d_{\ell}} \\ &= [B]_{\ell,n+1} - [B]_{\ell,f_1} x_{f_1} - [B]_{\ell,f_2} x_{f_2} - [B]_{\ell,f_3} x_{f_3} - \dots - [B]_{\ell,f_{n-r}} x_{f_{n-r}} \\ &= [\mathbf{c}]_{d_{\ell}} + x_{f_1} [\mathbf{u}_1]_{d_{\ell}} + x_{f_2} [\mathbf{u}_2]_{d_{\ell}} + x_{f_3} [\mathbf{u}_3]_{d_{\ell}} + \dots + x_{f_{n-r}} [\mathbf{u}_{n-r}]_{d_{\ell}} \\ &= [\mathbf{c} + x_{f_1} \mathbf{u}_1 + x_{f_2} \mathbf{u}_2 + x_{f_3} \mathbf{u}_3 + \dots + x_{f_{n-r}} \mathbf{u}_{n-r}]_{d_{\ell}} \end{aligned}$$

This tells us that the entries of the solution vector \mathbf{x} corresponding to dependent variables (indices in D), are equal to those of a vector in the set S. We still need to check the other entries of the solution vector \mathbf{x} corresponding to the free variables (indices in F) to see if they are equal to the entries of the same vector in the set S. To this end, suppose $i \in F$ and $i = f_i$. Then

$$\begin{bmatrix} \mathbf{x} \end{bmatrix}_i = x_i = x_{f_j} \\ = 0 + 0x_{f_1} + 0x_{f_2} + 0x_{f_3} + \dots + 0x_{f_{j-1}} + 1x_{f_j} + 0x_{f_{j+1}} + \dots + 0x_{f_{n-r}} \end{bmatrix}$$

$$= [\mathbf{c}]_i + x_{f_1} [\mathbf{u}_1]_i + x_{f_2} [\mathbf{u}_2]_i + x_{f_3} [\mathbf{u}_3]_i + \dots + x_{f_j} [\mathbf{u}_j]_i + \dots + x_{f_{n-r}} [\mathbf{u}_{n-r}]_i$$

= $[\mathbf{c} + x_{f_1} \mathbf{u}_1 + x_{f_2} \mathbf{u}_2 + \dots + x_{f_{n-r}} \mathbf{u}_{n-r}]_i$

So entries of \mathbf{x} and $\mathbf{c} + x_{f_1}\mathbf{u}_1 + x_{f_2}\mathbf{u}_2 + \cdots + x_{f_{n-r}}\mathbf{u}_{n-r}$ are equal and therefore by Definition CVE [199] they are equal vectors. Since $x_{f_1}, x_{f_2}, x_{f_3}, \ldots, x_{f_{n-r}}$ are scalars, this shows us that \mathbf{x} qualifies for membership in S. So the set S contains all of the solutions to the system.

Note that both halves of the proof of Theorem VFSLS [244] indicate that $\alpha_i = [\mathbf{x}]_{f_i}$. In other words, the arbitrary scalars, α_i , in the description of the set S actually have more meaning — they are the values of the free variables $[\mathbf{x}]_{f_i}$, $1 \leq i \leq n - r$. So we will often exploit this observation in our descriptions of solution sets.

Theorem VFSLS [244] formalizes what happened in the three steps of Example VFSAD [232]. The theorem will be useful in proving other theorems, and it it is useful since it tells us an exact procedure for simply describing an infinite solution set. We could program a computer to implement it, once we have the augmented matrix row-reduced and have checked that the system is consistent. By Knuth's definition, this completes our conversion of linear equation solving from art into science. Notice that it even applies (but is overkill) in the case of a unique solution. However, as a practical matter, I prefer the three-step process of Example VFSAD [232] when I need to describe an infinite solution set. So let's practice some more, but with a bigger example.

Example VFSAI

Vector form of solutions for Archetype I

Archetype I [1714] is a linear system of m = 4 equations in n = 7 variables. Row-reducing

the augmented matrix yields

$$\begin{bmatrix} 1 & 4 & 0 & 0 & 2 & 1 & -3 & 4 \\ 0 & 0 & 1 & 0 & 1 & -3 & 5 & 2 \\ 0 & 0 & 0 & 1 & 2 & -6 & 6 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and we see r = 3 nonzero rows. The columns with leading 1's are $D = \{1, 3, 4\}$ so the r dependent variables are x_1, x_3, x_4 . The columns without leading 1's are $F = \{2, 5, 6, 7, 8\}$, so the n - r = 4 free variables are x_2, x_5, x_6, x_7 .

Step 1. Write the vector of variables (\mathbf{x}) as a fixed vector (\mathbf{c}) , plus a linear combination of n - r = 4 vectors $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4)$, using the free variables as the scalars.

Step 2. For each free variable, use 0's and 1's to ensure equality for the corresponding entry of the vectors. Take note of the pattern of 0's and 1's at this stage, because this is the best look you'll have at it. We'll state an important theorem in the next section and the proof

will essentially rely on this observation.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Step 3. For each dependent variable, use the augmented matrix to formulate an equation expressing the dependent variable as a constant plus multiples of the free variables. Convert this equation into entries of the vectors that ensure equality for each dependent variable, one at a time.

$$\mathbf{x}_{1} = 4 - 4x_{2} - 2x_{5} - 1x_{6} + 3x_{7} \quad \Rightarrow$$

$$\mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{2} \begin{bmatrix} -4 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{5} \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{6} \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\$$

$$\begin{aligned} \mathbf{x}_{3} &= 2 + 0x_{2} - x_{5} + 3x_{6} - 5x_{7} \quad \Rightarrow \\ \mathbf{x} &= \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{2} \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{5} \begin{bmatrix} -2 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{6} \begin{bmatrix} -1 \\ 0 \\ 3 \\ 0 \\ 1 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ -5 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\ \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{4} &= 1 + 0x_{2} - 2x_{5} + 6x_{6} - 6x_{7} \quad \Rightarrow \\ \mathbf{x}_{5} &= \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \\ x_{6} \\ x_{7} \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{2} \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{5} \begin{bmatrix} -2 \\ 0 \\ -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_{6} \begin{bmatrix} -1 \\ 0 \\ 3 \\ 6 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_{7} \begin{bmatrix} 3 \\ 0 \\ -5 \\ -6 \\ 0 \\ 0 \\ 1 \end{bmatrix} \\$$

We can now use this final expression to quickly build solutions to the system. You might try to recreate each of the solutions listed in the write-up for Archetype I [1714]. (Hint: look at the values of the free variables in each solution, and notice that the vector \mathbf{c} has 0's in these

locations.)

Even better, we have a description of the infinite solution set, based on just 5 vectors, which we combine in linear combinations to produce solutions.

Whenever we discuss Archetype I [1714] you know that's your cue to go work through Archetype J [1723] by yourself. Remember to take note of the 0/1 pattern at the conclusion of Step 2. Have fun — we won't go anywhere while you're away.

This technique is so important, that we'll do one more example. However, an important distinction will be that this system is homogeneous.

Example VFSAL

Vector form of solutions for Archetype L

Archetype L [1742] is presented simply as the 5×5 matrix

$$L = \begin{bmatrix} -2 & -1 & -2 & -4 & 4 \\ -6 & -5 & -4 & -4 & 6 \\ 10 & 7 & 7 & 10 & -13 \\ -7 & -5 & -6 & -9 & 10 \\ -4 & -3 & -4 & -6 & 6 \end{bmatrix}$$

We'll interpret it here as the coefficient matrix of a homogeneous system and reference this matrix as L. So we are solving the homogeneous system $\mathcal{LS}(L, \mathbf{0})$ having m = 5 equations in n = 5 variables. If we built the augmented matrix, we would add a sixth column to L containing all zeros. As we did row operations, this sixth column would remain all zeros. So instead we will row-reduce the coefficient matrix, and mentally remember the missing sixth

column of zeros. This row-reduced matrix is

and we see r = 3 nonzero rows. The columns with leading 1's are $D = \{1, 2, 3\}$ so the r dependent variables are x_1, x_2, x_3 . The columns without leading 1's are $F = \{4, 5\}$, so the n - r = 2 free variables are x_4, x_5 . Notice that if we had included the all-zero vector of constants to form the augmented matrix for the system, then the index 6 would have appeared in the set F, and subsequently would have been ignored when listing the free variables.

Step 1. Write the vector of variables (\mathbf{x}) as a fixed vector (\mathbf{c}) , plus a linear combination of n - r = 2 vectors $(\mathbf{u}_1, \mathbf{u}_2)$, using the free variables as the scalars.

Step 2. For each free variable, use 0's and 1's to ensure equality for the corresponding entry of the vectors. Take note of the pattern of 0's and 1's at this stage, even if it is not as

illuminating as in other examples.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} \\ 0 \\ 1 \end{bmatrix}$$

Step 3. For each dependent variable, use the augmented matrix to formulate an equation expressing the dependent variable as a constant plus multiples of the free variables. Don't forget about the "missing" sixth column being full of zeros. Convert this equation into entries of the vectors that ensure equality for each dependent variable, one at a time.

$$x_{1} = 0 - 1x_{4} + 2x_{5} \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_{5} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$
$$x_{2} = 0 + 2x_{4} - 2x_{5} \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + x_{5} \begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

$$x_{3} = 0 - 2x_{4} + 1x_{5} \qquad \Rightarrow \qquad \mathbf{x} = \begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \\ x_{4} \\ x_{5} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_{4} \begin{bmatrix} -1 \\ 2 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_{5} \begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

The vector **c** will always have 0's in the entries corresponding to free variables. However, since we are solving a homogeneous system, the row-reduced augmented matrix has zeros in column n + 1 = 6, and hence *all* the entries of **c** are zero. So we can write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \mathbf{0} + x_4 \begin{bmatrix} -1 \\ 2 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \\ 1 \end{bmatrix} = x_4 \begin{bmatrix} -1 \\ 2 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

It will always happen that the solutions to a homogeneous system has $\mathbf{c} = \mathbf{0}$ (even in the case of a unique solution?). So our expression for the solutions is a bit more pleasing. In this example it says that the solutions are *all possible* linear combinations of the two vectors

$$\mathbf{u}_{1} = \begin{bmatrix} -1\\ 2\\ -2\\ 1\\ 0 \end{bmatrix} \text{ and } \mathbf{u}_{2} = \begin{bmatrix} 2\\ -2\\ 1\\ 0\\ 1 \end{bmatrix}, \text{ with no mention of any fixed vector entering into the linear combination.}$$

This observation will motivate our next section and the main definition of that section, and after that we will conclude the section by formalizing this situation. \square

Subsection PSHS Particular Solutions, Homogeneous Solutions

The next theorem tells us that in order to find all of the solutions to a linear system of equations, it is sufficient to find just one solution, and then find all of the solutions to the corresponding homogeneous system. This explains part of our interest in the null space, the set of all solutions to a homogeneous system.

Theorem **PSPHS**

Particular Solution Plus Homogeneous Solutions

Suppose that \mathbf{w} is one solution to the linear system of equations $\mathcal{LS}(A, b)$. Then \mathbf{y} is a solution to $\mathcal{LS}(A, b)$ if and only if $\mathbf{y} = \mathbf{w} + \mathbf{z}$ for some vector $\mathbf{z} \in \mathcal{N}(A)$.

Proof Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n$ be the columns of the coefficient matrix A. (\Leftarrow) Suppose $\mathbf{y} = \mathbf{w} + \mathbf{z}$ and $\mathbf{z} \in \mathcal{N}(A)$. Then

$$\mathbf{b} = [\mathbf{w}]_1 \mathbf{A}_1 + [\mathbf{w}]_2 \mathbf{A}_2 + [\mathbf{w}]_3 \mathbf{A}_3 + \dots + [\mathbf{w}]_n \mathbf{A}_n$$
 Theorem SLSLC [229]
$$= [\mathbf{w}]_1 \mathbf{A}_1 + [\mathbf{w}]_2 \mathbf{A}_2 + [\mathbf{w}]_3 \mathbf{A}_3 + \dots + [\mathbf{w}]_n \mathbf{A}_n + \mathbf{0}$$
 Property ZC [206]
$$= [\mathbf{w}]_1 \mathbf{A}_1 + [\mathbf{w}]_2 \mathbf{A}_2 + [\mathbf{w}]_3 \mathbf{A}_3 + \dots + [\mathbf{w}]_n \mathbf{A}_n$$
 Theorem SLSLC [229]

$$+ [\mathbf{z}]_{1} \mathbf{A}_{1} + [\mathbf{z}]_{2} \mathbf{A}_{2} + [\mathbf{z}]_{3} \mathbf{A}_{3} + \dots + [\mathbf{z}]_{n} \mathbf{A}_{n}$$

$$= ([\mathbf{w}]_{1} + [\mathbf{z}]_{1}) \mathbf{A}_{1} + ([\mathbf{w}]_{2} + [\mathbf{z}]_{2}) \mathbf{A}_{2} + \dots + ([\mathbf{w}]_{n} + [\mathbf{z}]_{n}) \mathbf{A}_{n}$$

$$= [\mathbf{w} + \mathbf{z}]_{1} \mathbf{A}_{1} + [\mathbf{w} + \mathbf{z}]_{2} \mathbf{A}_{2} + [\mathbf{w} + \mathbf{z}]_{3} \mathbf{A}_{3} + \dots + [\mathbf{w} + \mathbf{z}]_{n} \mathbf{A}_{n}$$

$$= [\mathbf{y}]_{1} \mathbf{A}_{1} + [\mathbf{y}]_{2} \mathbf{A}_{2} + [\mathbf{y}]_{3} \mathbf{A}_{3} + \dots + [\mathbf{y}]_{n} \mathbf{A}_{n}$$

$$Definition CVA [201]$$

$$= [\mathbf{y}]_{1} \mathbf{A}_{1} + [\mathbf{y}]_{2} \mathbf{A}_{2} + [\mathbf{y}]_{3} \mathbf{A}_{3} + \dots + [\mathbf{y}]_{n} \mathbf{A}_{n}$$

$$Definition of \mathbf{y}$$

Applying Theorem SLSLC [229] we see that the vector \mathbf{y} is a solution to $\mathcal{LS}(A, \mathbf{b})$. (\Rightarrow) Suppose \mathbf{y} is a solution to $\mathcal{LS}(A, b)$. Then

$$\begin{aligned} \mathbf{0} &= \mathbf{b} - \mathbf{b} \\ &= [\mathbf{y}]_1 \mathbf{A}_1 + [\mathbf{y}]_2 \mathbf{A}_2 + [\mathbf{y}]_3 \mathbf{A}_3 + \dots + [\mathbf{y}]_n \mathbf{A}_n & \text{Theorem SLSLC [229]} \\ &- ([\mathbf{w}]_1 \mathbf{A}_1 + [\mathbf{w}]_2 \mathbf{A}_2 + [\mathbf{w}]_3 \mathbf{A}_3 + \dots + [\mathbf{w}]_n \mathbf{A}_n) \\ &= ([\mathbf{y}]_1 - [\mathbf{w}]_1) \mathbf{A}_1 + ([\mathbf{y}]_2 - [\mathbf{w}]_2) \mathbf{A}_2 + \dots + ([\mathbf{y}]_n - [\mathbf{w}]_n) \mathbf{A}_n & \text{Theorem VSPCV [205]} \\ &= [\mathbf{y} - \mathbf{w}]_1 \mathbf{A}_1 + [\mathbf{y} - \mathbf{w}]_2 \mathbf{A}_2 + [\mathbf{y} - \mathbf{w}]_3 \mathbf{A}_3 + \dots + [\mathbf{y} - \mathbf{w}]_n \mathbf{A}_n & \text{Definition CVA [201]} \end{aligned}$$

By Theorem SLSLC [229] we see that the vector $\mathbf{y} - \mathbf{w}$ is a solution to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ and by Definition NSM [152], $\mathbf{y} - \mathbf{w} \in \mathcal{N}(A)$. In other words, $\mathbf{y} - \mathbf{w} = \mathbf{z}$ for some vector $\mathbf{z} \in \mathcal{N}(A)$. Rewritten, this is $\mathbf{y} = \mathbf{w} + \mathbf{z}$, as desired.

After proving Theorem NMUS [180] we commented (insufficiently) on the negation of one half of the theorem. Nonsingular coefficient matrices lead to unique solutions for every choice of the vector of constants. What does this say about singular matrices? A singular matrix A has a nontrivial null space (Theorem NMTNS [179]). For a given vector of constants, **b**, the system $\mathcal{LS}(A, b)$ could be inconsistent, meaning there are no solutions. But if there is at least one solution (\mathbf{w}), then Theorem PSPHS [256] tells us there will be infinitely many solutions because of the role of the infinite null space for a singular matrix. So a system of equations with a singular coefficient matrix *never* has a unique solution. Either there are no solutions, or infinitely many solutions, depending on the choice of the vector of constants (\mathbf{b}).

Example PSHS

Particular solutions, homogeneous solutions, Archetype D

Archetype D [1670] is a consistent system of equations with a nontrivial null space. Let A denote the coefficient matrix of this system. The write-up for this system begins with three solutions,

$$\mathbf{y}_1 = \begin{bmatrix} 0\\1\\2\\1 \end{bmatrix} \qquad \qquad \mathbf{y}_2 = \begin{bmatrix} 4\\0\\0\\0 \end{bmatrix} \qquad \qquad \mathbf{y}_3 = \begin{bmatrix} 7\\8\\1\\3 \end{bmatrix}$$

We will choose to have \mathbf{y}_1 play the role of \mathbf{w} in the statement of Theorem PSPHS [256], any one of the three vectors listed here (or others) could have been chosen. To illustrate the theorem, we should be able to write each of these three solutions as the vector \mathbf{w} plus a solution to the corresponding homogeneous system of equations. Since $\mathbf{0}$ is always a solution to a homogeneous system we can easily write

$$\mathbf{y}_1 = \mathbf{w} = \mathbf{w} + \mathbf{0}.$$

The vectors \mathbf{y}_2 and \mathbf{y}_3 will require a bit more effort. Solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ are exactly the elements of the null space of the coefficient matrix, which by an application of Theorem VFSLS [244] is

$$\mathcal{N}(A) = \left\{ x_3 \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix} + x_4 \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix} \middle| x_3, x_4 \in \mathbb{C} \right\}$$

Then

$$\mathbf{y}_{2} = \begin{bmatrix} 4\\0\\0\\0 \end{bmatrix} = \begin{bmatrix} 0\\1\\2\\1 \end{bmatrix} + \begin{bmatrix} 4\\-1\\-2\\-1 \end{bmatrix} = \begin{bmatrix} 0\\1\\2\\1 \end{bmatrix} + \left((-2)\begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix} + (-1)\begin{bmatrix} 2\\3\\0\\1 \end{bmatrix} \right) = \mathbf{w} + \mathbf{z}_{2}$$

where

$$\mathbf{z}_{2} = \begin{bmatrix} 4\\-1\\-2\\-1 \end{bmatrix} = (-2) \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix} + (-1) \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix}$$

is obviously a solution of the homogeneous system since it is written as a linear combination of the vectors describing the null space of the coefficient matrix (or as a check, you could just evaluate the equations in the homogeneous system with \mathbf{z}_2). Again

$$\mathbf{y}_{3} = \begin{bmatrix} 7\\8\\1\\3 \end{bmatrix} = \begin{bmatrix} 0\\1\\2\\1 \end{bmatrix} + \begin{bmatrix} 7\\7\\-1\\2\\1 \end{bmatrix} = \begin{bmatrix} 0\\1\\2\\1 \end{bmatrix} + \left((-1) \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix} + 2 \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix} \right) = \mathbf{w} + \mathbf{z}_{3}$$

where

$$\mathbf{z}_{3} = \begin{bmatrix} 7\\7\\-1\\2 \end{bmatrix} = (-1) \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix} + 2 \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix}$$

is obviously a solution of the homogeneous system since it is written as a linear combination of the vectors describing the null space of the coefficient matrix (or as a check, you could just evaluate the equations in the homogeneous system with \mathbf{z}_2).

Here's another view of this theorem, in the context of this example. Grab two new solutions of the original system of equations, say

$$\mathbf{y}_4 = \begin{bmatrix} 11\\0\\-3\\-1 \end{bmatrix} \qquad \qquad \mathbf{y}_5 = \begin{bmatrix} -4\\2\\4\\2 \end{bmatrix}$$

and form their difference,

$$\mathbf{u} = \begin{bmatrix} 11\\0\\-3\\-1 \end{bmatrix} - \begin{bmatrix} -4\\2\\4\\2 \end{bmatrix} = \begin{bmatrix} 15\\-2\\-7\\-3 \end{bmatrix}$$

It is no accident that \mathbf{u} is a solution to the homogeneous system (check this!). In other words, the difference between any two solutions to a linear system of equations is an element of the null space of the coefficient matrix. This is an equivalent way to state Theorem PSPHS [256]. (See Exercise MM.T50 [494]).

The ideas of this subsection will appear again in Chapter LT [1071] when we discuss pre-images of linear transformations (Definition PI [1101]).

Subsection READ Reading Questions

1. Earlier, a reading question asked you to solve the system of equations

$$2x_1 + 3x_2 - x_3 = 0$$

$$x_1 + 2x_2 + x_3 = 3$$

$$x_1 + 3x_2 + 3x_3 = 7$$

Use a linear combination to rewrite this system of equations as a vector equality.

2. Find a linear combination of the vectors

$$S = \left\{ \begin{bmatrix} 1\\3\\-1 \end{bmatrix}, \begin{bmatrix} 2\\0\\4 \end{bmatrix}, \begin{bmatrix} -1\\3\\-5 \end{bmatrix} \right\}$$
 that equals the vector $\begin{bmatrix} 1\\-9\\11 \end{bmatrix}$.

3. The matrix below is the augmented matrix of a system of equations, row-reduced to reduced row-echelon form. Write the vector form of the solutions to the system.

$$\begin{bmatrix} 1 & 3 & 0 & 6 & 0 & 9 \\ 0 & 0 & 1 & -2 & 0 & -8 \\ 0 & 0 & 0 & 0 & 1 & 3 \end{bmatrix}$$

Subsection EXC Exercises

C21 Consider each archetype that is a system of equations. For individual solutions listed (both for the original system and the corresponding homogeneous system) express the vector of constants as a linear combination of the columns of the coefficient matrix, as guaranteed by Theorem SLSLC [229]. Verify this equality by computing the linear combination. For systems with no solutions, recognize that it is then impossible to write the vector of constants as a linear combination of the coefficient matrix. Note too, for homogeneous systems, that the solutions give rise to linear combinations that equal the zero vector.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670] Archetype E [1678] Archetype F [1686] Archetype G [1697] Archetype H [1705] Archetype I [1714] Archetype J [1723] Contributed by Robert Beezer Solution [267]

C22 Consider each archetype that is a system of equations. Write elements of the solution set in vector form, as guaranteed by Theorem VFSLS [244].

Archetype A [1642]Archetype B [1652]Archetype C [1662]Archetype D [1670]Archetype E [1678]Archetype F [1686]Archetype G [1697]Archetype H [1705]Archetype I [1714]Archetype J [1723]

Contributed by Robert Beezer Solution [267]

C40 Find the vector form of the solutions to the system of equations below.

$$2x_1 - 4x_2 + 3x_3 + x_5 = 6$$
$$x_1 - 2x_2 - 2x_3 + 14x_4 - 4x_5 = 15$$
$$x_1 - 2x_2 + x_3 + 2x_4 + x_5 = -1$$
$$-2x_1 + 4x_2 - 12x_4 + x_5 = -7$$

Contributed by Robert Beezer Solution [267]

C41 Find the vector form of the solutions to the system of equations below.

$$\begin{aligned} -2x_1 - 1x_2 - 8x_3 + 8x_4 + 4x_5 - 9x_6 - 1x_7 - 1x_8 - 18x_9 &= 3\\ 3x_1 - 2x_2 + 5x_3 + 2x_4 - 2x_5 - 5x_6 + 1x_7 + 2x_8 + 15x_9 &= 10\\ 4x_1 - 2x_2 + 8x_3 + 2x_5 - 14x_6 - 2x_8 + 2x_9 &= 36\\ -1x_1 + 2x_2 + 1x_3 - 6x_4 + 7x_6 - 1x_7 - 3x_9 &= -8\\ 3x_1 + 2x_2 + 13x_3 - 14x_4 - 1x_5 + 5x_6 - 1x_8 + 12x_9 &= 15\\ -2x_1 + 2x_2 - 2x_3 - 4x_4 + 1x_5 + 6x_6 - 2x_7 - 2x_8 - 15x_9 &= -7 \end{aligned}$$

Contributed by Robert Beezer Solution [268] M10 Example TLC [222] asks if the vector

$$\mathbf{w} = \begin{bmatrix} 13\\15\\5\\-17\\2\\25 \end{bmatrix}$$

can be written as a linear combination of the four vectors

$$\mathbf{u}_{1} = \begin{bmatrix} 2\\4\\-3\\1\\2\\9 \end{bmatrix} \qquad \mathbf{u}_{2} = \begin{bmatrix} 6\\3\\0\\-2\\1\\4 \end{bmatrix} \qquad \mathbf{u}_{3} = \begin{bmatrix} -5\\2\\1\\1\\-3\\0 \end{bmatrix} \qquad \mathbf{u}_{4} = \begin{bmatrix} 3\\2\\-5\\7\\1\\3 \end{bmatrix}$$

Can it? Can any vector in \mathbb{C}^6 be written as a linear combination of the four vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$? Contributed by Robert Beezer Solution [269]

M11 At the end of Example VFS [237], the vector \mathbf{w} is claimed to be a solution to the linear system under discussion. Verify that \mathbf{w} really is a solution. Then determine the four scalars that express \mathbf{w} as a linear combination of \mathbf{c} , \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 . Contributed by Robert Beezer Solution [269]

Subsection SOL Solutions

C21 Contributed by Robert Beezer Statement [263] Solutions for Archetype A [1642] and Archetype B [1652] are described carefully in Example AALC [227] and Example ABLC [225].

C22 Contributed by Robert Beezer Statement [264]

Solutions for Archetype D [1670] and Archetype I [1714] are described carefully in Example VFSAD [232] and Example VFSAI [248]. The technique described in these examples is probably more useful than carefully deciphering the notation of Theorem VFSLS [244]. The solution for each archetype is contained in its description. So now you can check-off the box for that item.

C40 Contributed by Robert Beezer Statement [264]

Row-reduce the augmented matrix representing this system, to find

$$\begin{bmatrix} 1 & -2 & 0 & 6 & 0 & 1 \\ 0 & 0 & 1 & -4 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The system is consistent (no leading one in column 6, Theorem RCLS [120]). x_2 and x_4

are the free variables. Now apply Theorem VFSLS [244] directly, or follow the three-step process of Example VFS [237], Example VFSAD [232], Example VFSAI [248], or Example VFSAL [252] to obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -6 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix}$$

C41 Contributed by Robert Beezer Statement [265] Row-reduce the augmented matrix representing this system, to find

The system is consistent (no leading one in column 10, Theorem RCLS [120]). $F = \{3, 4, 6, 9, 10\}$, so the free variables are x_3, x_4, x_6 and x_9 . Now apply Theorem VFSLS [244] directly, or follow the three-step process of Example VFS [237], Example VFSAD

[232], Example VFSAI [248], or Example VFSAL [252] to obtain the solution set

M10 Contributed by Robert Beezer Statement [265]

No, it is not possible to create \mathbf{w} as a linear combination of the four vectors \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 , \mathbf{u}_4 . By creating the desired linear combination with unknowns as scalars, Theorem SLSLC [229] provides a system of equations that has no solution. This one computation is enough to show us that it is not possible to create all the vectors of \mathbb{C}^6 through linear combinations of the four vectors \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 , \mathbf{u}_4 .

M11 Contributed by Robert Beezer Statement [266]

The coefficient of **c** is 1. The coefficients of \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 lie in the third, fourth and seventh entries of **w**. Can you see why? (Hint: $F = \{3, 4, 7, 8\}$, so the free variables are x_3 , x_4 and x_7 .)

Section SS Spanning Sets

In this section we will describe a compact way to indicate the elements of an infinite set of vectors, making use of linear combinations. This will give us a convenient way to describe the elements of a set of solutions to a linear system, or the elements of the null space of a matrix, or many other sets of vectors.

Subsection SSV Span of a Set of Vectors

In Example VFSAL [252] we saw the solution set of a homogeneous system described as all possible linear combinations of two particular vectors. This happens to be a useful way to construct or describe infinite sets of vectors, so we encapsulate this idea in a definition.

Definition SSCV Span of a Set of Column Vectors Given a set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_p}$, their span, $\langle S \rangle$, is the set of all possible linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_p$. Symbolically,

$$\langle S \rangle = \left\{ \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_p \mathbf{u}_p \middle| \alpha_i \in \mathbb{C}, \ 1 \le i \le p \right\}$$
$$= \left\{ \sum_{i=1}^p \alpha_i \mathbf{u}_i \middle| \alpha_i \in \mathbb{C}, \ 1 \le i \le p \right\}$$

(This definition contains Notation SSV.)

The span is just a set of vectors, though in all but one situation it is an infinite set. (Just when is it not infinite?) So we start with a finite collection of vectors S (p of them to be precise), and use this finite set to describe an infinite set of vectors, $\langle S \rangle$. Confusing the finite set S with the *infinite* set $\langle S \rangle$ is one of the most pervasive problems in understanding introductory linear algebra. We will see this construction repeatedly, so let's work through some examples to get comfortable with it. The most obvious question about a set is if a particular item of the correct type is in the set, or not.

Example ABS

A basic span

Consider the set of 5 vectors, S, from \mathbb{C}^4

$$S = \left\{ \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix}, \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix}, \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix}, \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} \right\}$$

Δ

and consider the infinite set of vectors $\langle S \rangle$ formed from all possible linear combinations of the elements of S. Here are four vectors we definitely know are elements of $\langle S \rangle$, since we will construct them in accordance with Definition SSCV [270],

$$\mathbf{w} = (2) \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix} + (1) \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix} + (-1) \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix} + (2) \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix} + (3) \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} = \begin{bmatrix} -4\\2\\2\\8\\10 \end{bmatrix}$$
$$\mathbf{x} = (5) \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix} + (-6) \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix} + (-3) \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix} + (4) \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix} + (2) \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} = \begin{bmatrix} -26\\-6\\2\\34 \end{bmatrix}$$
$$\mathbf{y} = (1) \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix} + (0) \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix} + (1) \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix} + (0) \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix} + (1) \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} = \begin{bmatrix} 7\\4\\17\\-4 \end{bmatrix}$$
$$\mathbf{z} = (0) \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix} + (0) \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix} + (0) \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix} + (0) \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix} + (0) \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} = \begin{bmatrix} 7\\4\\17\\-4 \end{bmatrix}$$

The purpose of a set is to collect objects with some common property, and to exclude objects without that property. So the most fundamental question about a set is if a given object is

an element of the set or not. Let's learn more about $\langle S \rangle$ by investigating which vectors are elements of the set, and which are not.

First, is $\mathbf{u} = \begin{bmatrix} -15 \\ -6 \\ 19 \\ 5 \end{bmatrix}$ an element of $\langle S \rangle$? We are asking if there are scalars $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$

such that

$$\alpha_{1} \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix} + \alpha_{2} \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix} + \alpha_{3} \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix} + \alpha_{4} \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix} + \alpha_{5} \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} = \mathbf{u} = \begin{bmatrix} -15\\-6\\19\\5 \end{bmatrix}$$

Applying Theorem SLSLC [229] we recognize the search for these scalars as a solution to a linear system of equations with augmented matrix

$$\begin{bmatrix} 1 & 2 & 7 & 1 & -1 & -15 \\ 1 & 1 & 3 & 1 & 0 & -6 \\ 3 & 2 & 5 & -1 & 9 & 19 \\ 1 & -1 & -5 & 2 & 0 & 5 \end{bmatrix}$$

which row-reduces to

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 3 & 10 \\ 0 & 1 & 4 & 0 & -1 & -9 \\ 0 & 0 & 0 & 1 & -2 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

At this point, we see that the system is consistent (Theorem RCLS [120]), so we know there is a solution for the five scalars $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$. This is enough evidence for us to say that $\mathbf{u} \in \langle S \rangle$. If we wished further evidence, we could compute an actual solution, say

$$\alpha_1 = 2 \qquad \qquad \alpha_2 = 1 \qquad \qquad \alpha_3 = -2 \qquad \qquad \alpha_4 = -3 \qquad \qquad \alpha_5 = 2$$

This particular solution allows us to write

$$(2)\begin{bmatrix}1\\1\\3\\1\end{bmatrix} + (1)\begin{bmatrix}2\\1\\2\\-1\end{bmatrix} + (-2)\begin{bmatrix}7\\3\\5\\-5\end{bmatrix} + (-3)\begin{bmatrix}1\\1\\-1\\2\end{bmatrix} + (2)\begin{bmatrix}-1\\0\\9\\0\end{bmatrix} = \mathbf{u} = \begin{bmatrix}-15\\-6\\19\\5\end{bmatrix}$$

making it even more obvious that $\mathbf{u} \in \langle S \rangle$.

Lets do it again. Is
$$\mathbf{v} = \begin{bmatrix} 3\\1\\2\\-1 \end{bmatrix}$$
 an element of $\langle S \rangle$? We are asking if there are scalars

 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ such that

$$\alpha_{1} \begin{bmatrix} 1\\1\\3\\1 \end{bmatrix} + \alpha_{2} \begin{bmatrix} 2\\1\\2\\-1 \end{bmatrix} + \alpha_{3} \begin{bmatrix} 7\\3\\5\\-5 \end{bmatrix} + \alpha_{4} \begin{bmatrix} 1\\1\\-1\\2 \end{bmatrix} + \alpha_{5} \begin{bmatrix} -1\\0\\9\\0 \end{bmatrix} = \mathbf{v} = \begin{bmatrix} 3\\1\\2\\-1 \end{bmatrix}$$

Applying Theorem SLSLC [229] we recognize the search for these scalars as a solution to a linear system of equations with augmented matrix

[1	2	7	1	-1	3
1	1	3	1	0	1
3	2	5	-1	9	2
1	-1	-5	2	$-1 \\ 0 \\ 9 \\ 0$	$-1_{_{_{_{_{_{}}}}}}$

which row-reduces to

$\lceil 1 \rceil$	0	-1	0	3	0]
0	1	4	0	-1	0
$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	0	0	1	-2	0
0	0	0	0	0	1

At this point, we see that the system is inconsistent by Theorem RCLS [120], so we know there *is not* a solution for the five scalars α_1 , α_2 , α_3 , α_4 , α_5 . This is enough evidence for us to say that $\mathbf{v} \notin \langle S \rangle$. End of story.

Example SCAA Span of the columns of Archetype A Begin with the finite set of three vectors of size 3

$$S = \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 1\\2\\1 \end{bmatrix}, \, \begin{bmatrix} -1\\1\\1 \end{bmatrix}, \, \begin{bmatrix} 2\\1\\0 \end{bmatrix} \right\}$$

and consider the infinite set $\langle S \rangle$. The vectors of S could have been chosen to be anything, but for reasons that will become clear later, we have chosen the three columns of the coefficient matrix in Archetype A [1642]. First, as an example, note that

$$\mathbf{v} = (5) \begin{bmatrix} 1\\2\\1 \end{bmatrix} + (-3) \begin{bmatrix} -1\\1\\1 \end{bmatrix} + (7) \begin{bmatrix} 2\\1\\0 \end{bmatrix} = \begin{bmatrix} 22\\14\\2 \end{bmatrix}$$

is in $\langle S \rangle$, since it is a linear combination of \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 . We write this succinctly as $\mathbf{v} \in \langle S \rangle$. There is nothing magical about the scalars $\alpha_1 = 5$, $\alpha_2 = -3$, $\alpha_3 = 7$, they could have been chosen to be anything. So repeat this part of the example yourself, using different values of α_1 , α_2 , α_3 . What happens if you choose all three scalars to be zero?

So we know how to quickly construct sample elements of the set $\langle S \rangle$. A slightly different question arises when you are handed a vector of the correct size and asked if it is an element

of
$$\langle S \rangle$$
. For example, is $\mathbf{w} = \begin{bmatrix} 1\\ 8\\ 5 \end{bmatrix}$ in $\langle S \rangle$? More succinctly, $\mathbf{w} \in \langle S \rangle$?

To answer this question, we will look for scalars α_1 , α_2 , α_3 so that

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 = \mathbf{w}$$

By Theorem SLSLC [229] solutions to this vector equation are solutions to the system of equations

$$\alpha_1 - \alpha_2 + 2\alpha_3 = 1$$

$$2\alpha_1 + \alpha_2 + \alpha_3 = 8$$
$$\alpha_1 + \alpha_2 = 5$$

Building the augmented matrix for this linear system, and row-reducing, gives

$$\begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This system has infinitely many solutions (there's a free variable in x_3), but all we need is one solution vector. The solution,

$$\alpha_1 = 2 \qquad \qquad \alpha_2 = 3 \qquad \qquad \alpha_3 = 1$$

tells us that

$$(2)\mathbf{u}_1 + (3)\mathbf{u}_2 + (1)\mathbf{u}_3 = \mathbf{w}$$

so we are convinced that **w** really is in $\langle S \rangle$. Notice that there are an infinite number of ways to answer this question affirmatively. We could choose a different solution, this time choosing the free variable to be zero,

$$\alpha_1 = 3 \qquad \qquad \alpha_2 = 2 \qquad \qquad \alpha_3 = 0$$

shows us that

$$(3)\mathbf{u}_1 + (2)\mathbf{u}_2 + (0)\mathbf{u}_3 = \mathbf{w}$$

Version 2.99

Verifying the arithmetic in this second solution will make it obvious that \mathbf{w} is in this span. And of course, we now realize that there are an infinite number of ways to realize \mathbf{w} as element of $\langle S \rangle$. Let's ask the same type of question again, but this time with $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$, i.e.

is $\mathbf{y} \in \langle S \rangle$? So we'll look for scalars $\alpha_1, \alpha_2, \alpha_3$ so that

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 = \mathbf{y}$$

By Theorem SLSLC [229] solutions to this vector equation are the solutions to the system of equations

$$\alpha_1 - \alpha_2 + 2\alpha_3 = 2$$
$$2\alpha_1 + \alpha_2 + \alpha_3 = 4$$
$$\alpha_1 + \alpha_2 = 3$$

Building the augmented matrix for this linear system, and row-reducing, gives

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This system is inconsistent (there's a leading 1 in the last column, Theorem RCLS [120]), so there are no scalars α_1 , α_2 , α_3 that will create a linear combination of \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 that equals **y**. More precisely, $\mathbf{y} \notin \langle S \rangle$.

There are three things to observe in this example. (1) It is easy to construct vectors in $\langle S \rangle$. (2) It is possible that some vectors are in $\langle S \rangle$ (e.g. **w**), while others are not (e.g. **y**). (3) Deciding if a given vector is in $\langle S \rangle$ leads to solving a linear system of equations and asking if the system is consistent.

With a computer program in hand to solve systems of linear equations, could you create a program to decide if a vector was, or wasn't, in the span of a given set of vectors? Is this art or science?

This example was built on vectors from the columns of the coefficient matrix of Archetype A [1642]. Study the determination that $\mathbf{v} \in \langle S \rangle$ and see if you can connect it with some of the other properties of Archetype A [1642].

Having analyzed Archetype A [1642] in Example SCAA [275], we will of course subject Archetype B [1652] to a similar investigation.

Example SCAB

Span of the columns of Archetype B

Begin with the finite set of three vectors of size 3 that are the columns of the coefficient matrix in Archetype B [1652],

$$R = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3\} = \left\{ \begin{bmatrix} -7\\5\\1 \end{bmatrix}, \begin{bmatrix} -6\\5\\0 \end{bmatrix}, \begin{bmatrix} -12\\7\\4 \end{bmatrix} \right\}$$

and consider the infinite set $\langle R \rangle$. First, as an example, note that

$$\mathbf{x} = (2) \begin{bmatrix} -7\\5\\1 \end{bmatrix} + (4) \begin{bmatrix} -6\\5\\0 \end{bmatrix} + (-3) \begin{bmatrix} -12\\7\\4 \end{bmatrix} = \begin{bmatrix} -2\\9\\-10 \end{bmatrix}$$

is in $\langle R \rangle$, since it is a linear combination of \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 . In other words, $\mathbf{x} \in \langle R \rangle$. Try some different values of α_1 , α_2 , α_3 yourself, and see what vectors you can create as elements of $\langle R \rangle$.

Now ask if a given vector is an element of $\langle R \rangle$. For example, is $\mathbf{z} = \begin{bmatrix} -33\\ 24\\ 5 \end{bmatrix}$ in $\langle R \rangle$? Is

 $\mathbf{z} \in \langle R \rangle$?

To answer this question, we will look for scalars α_1 , α_2 , α_3 so that

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 = \mathbf{z}$$

By Theorem SLSLC [229] solutions to this vector equation are the solutions to the system of equations

$$-7\alpha_1 - 6\alpha_2 - 12\alpha_3 = -33$$
$$5\alpha_1 + 5\alpha_2 + 7\alpha_3 = 24$$
$$\alpha_1 + 4\alpha_3 = 5$$

Building the augmented matrix for this linear system, and row-reducing, gives

$$\begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

This system has a unique solution,

$$\alpha_1 = -3 \qquad \qquad \alpha_2 = 5 \qquad \qquad \alpha_3 = 2$$

telling us that

$$(-3)\mathbf{v}_1 + (5)\mathbf{v}_2 + (2)\mathbf{v}_3 = \mathbf{z}$$

so we are convinced that \mathbf{z} really is in $\langle R \rangle$. Notice that in this case we have only one way to answer the question affirmatively since the solution is unique.

Let's ask about another vector, say is
$$\mathbf{x} = \begin{bmatrix} -7\\ 8\\ -3 \end{bmatrix}$$
 in $\langle R \rangle$? Is $\mathbf{x} \in \langle R \rangle$?

We desire scalars α_1 , α_2 , α_3 so that

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 = \mathbf{x}$$

By Theorem SLSLC [229] solutions to this vector equation are the solutions to the system of equations

$$-7\alpha_1 - 6\alpha_2 - 12\alpha_3 = -7$$

$$5\alpha_1 + 5\alpha_2 + 7\alpha_3 = 8$$
$$\alpha_1 + 4\alpha_3 = -3$$

Building the augmented matrix for this linear system, and row-reducing, gives

1	0	0	1]
0	1	0	2
0	0	1	-1

This system has a unique solution,

$$\alpha_1 = 1 \qquad \qquad \alpha_2 = 2 \qquad \qquad \alpha_3 = -1$$

telling us that

$$(1)\mathbf{v}_1 + (2)\mathbf{v}_2 + (-1)\mathbf{v}_3 = \mathbf{x}$$

so we are convinced that \mathbf{x} really is in $\langle R \rangle$. Notice that in this case we again have only one way to answer the question affirmatively since the solution is again unique.

We could continue to test other vectors for membership in $\langle R \rangle$, but there is no point. A question about membership in $\langle R \rangle$ inevitably leads to a system of three equations in the three variables α_1 , α_2 , α_3 with a coefficient matrix whose columns are the vectors \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 . This particular coefficient matrix is nonsingular, so by Theorem NMUS [180], the system is guaranteed to have a solution. (This solution is unique, but that's not critical here.) So *no* matter which vector we might have chosen for \mathbf{z} , we would have been certain to discover that it was an element of $\langle R \rangle$. Stated differently, every vector of size 3 is in $\langle R \rangle$, or $\langle R \rangle = \mathbb{C}^3$. Compare this example with Example SCAA [275], and see if you can connect \mathbf{z} with some aspects of the write-up for Archetype B [1652].

Subsection SSNS Spanning Sets of Null Spaces

We saw in Example VFSAL [252] that when a system of equations is homogeneous the solution set can be expressed in the form described by Theorem VFSLS [244] where the vector **c** is the zero vector. We can essentially ignore this vector, so that the remainder of the typical expression for a solution looks like an arbitrary linear combination, where the scalars are the free variables and the vectors are $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_{n-r}$. Which sounds a lot like a span. This is the substance of the next theorem.

Theorem SSNS

Spanning Sets for Null Spaces

Suppose that A is an $m \times n$ matrix, and B is a row-equivalent matrix in reduced row-echelon form with r nonzero rows. Let $D = \{d_1, d_2, d_3, \ldots, d_r\}$ be the column indices where B has leading 1's (pivot columns) and $F = \{f_1, f_2, f_3, \ldots, f_{n-r}\}$ be the set of column indices where B does not have leading 1's. Construct the n-r vectors \mathbf{z}_i , $1 \leq j \leq n-r$ of size n as

$$[\mathbf{z}_j]_i = \begin{cases} 1 & \text{if } i \in F, \ i = f_j \\ 0 & \text{if } i \in F, \ i \neq f_j \\ -[B]_{k,f_i} & \text{if } i \in D, \ i = d_k \end{cases}$$

Then the null space of A is given by

$$\mathcal{N}(A) = \langle \{\mathbf{z}_1, \, \mathbf{z}_2, \, \mathbf{z}_3, \, \dots, \, \mathbf{z}_{n-r}\} \rangle$$

Proof Consider the homogeneous system with A as a coefficient matrix, $\mathcal{LS}(A, \mathbf{0})$. Its set of solutions, S, is by Definition NSM [152], the null space of A, $\mathcal{N}(A)$. Let B' denote the result of row-reducing the augmented matrix of this homogeneous system. Since the system is homogeneous, the final column of the augmented matrix will be all zeros, and after any number of row operations (Definition RO [59]), the column will still be all zeros. So B' has a final column that is totally zeros.

Now apply Theorem VFSLS [244] to B', after noting that our homogeneous system must be consistent (Theorem HSC [147]). The vector **c** has zeros for each entry that corresponds to an index in F. For entries that correspond to an index in D, the value is $-[B']_{k,n+1}$, but for B' any entry in the final column (index n + 1) is zero. So $\mathbf{c} = \mathbf{0}$. The vectors \mathbf{z}_j , $1 \leq j \leq n - r$ are identical to the vectors \mathbf{u}_j , $1 \leq j \leq n - r$ described in Theorem VFSLS [244]. Putting it all together and applying Definition SSCV [270] in the final step,

$$\mathcal{N}(A) = S$$

$$= \{ \mathbf{c} + \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_{n-r} \mathbf{u}_{n-r} | \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{n-r} \in \mathbb{C} \}$$

= $\{ \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_{n-r} \mathbf{u}_{n-r} | \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_{n-r} \in \mathbb{C} \}$
= $\langle \{ \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_{n-r} \} \rangle$

Example SSNS Spanning set of a null space

Find a set of vectors, S, so that the null space of the matrix A below is the span of S, that is, $\langle S \rangle = \mathcal{N}(A)$.

$$A = \begin{bmatrix} 1 & 3 & 3 & -1 & -5 \\ 2 & 5 & 7 & 1 & 1 \\ 1 & 1 & 5 & 1 & 5 \\ -1 & -4 & -2 & 0 & 4 \end{bmatrix}$$

The null space of A is the set of all solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. If we find the vector form of the solutions to this homogeneous system (Theorem VFSLS [244]) then the vectors \mathbf{u}_j , $1 \leq j \leq n-r$ in the linear combination are exactly the vectors \mathbf{z}_j , $1 \leq j \leq n-r$ described in Theorem SSNS [283]. So we can mimic Example VFSAL [252] to arrive at these vectors (rather than being a slave to the formulas in the statement of the theorem).

Begin by row-reducing A. The result is

$$\begin{bmatrix} 1 & 0 & 6 & 0 & 4 \\ 0 & 1 & -1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

With $D = \{1, 2, 4\}$ and $F = \{3, 5\}$ we recognize that x_3 and x_5 are free variables and we can express each nonzero row as an expression for the dependent variables x_1 , x_2 , x_4 (respectively) in the free variables x_3 and x_5 . With this we can write the vector form of a solution vector as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -6x_3 - 4x_5 \\ x_3 + 2x_5 \\ x_3 \\ -3x_5 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} -6 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -4 \\ 2 \\ 0 \\ -3 \\ 1 \end{bmatrix}$$

Then in the notation of Theorem SSNS [283],

$$\mathbf{z}_1 = \begin{bmatrix} -6\\1\\1\\0\\0 \end{bmatrix} \qquad \qquad \mathbf{z}_2 = \begin{bmatrix} -4\\2\\0\\-3\\1 \end{bmatrix}$$

and

$$\mathcal{N}(A) = \langle \{\mathbf{z}_1, \, \mathbf{z}_2\} \rangle = \left\langle \left\{ \begin{bmatrix} -6\\1\\1\\0\\0 \end{bmatrix}, \, \begin{bmatrix} -4\\2\\0\\-3\\1 \end{bmatrix} \right\} \right\rangle$$

Example NSDS Null space directly as a span

Let's express the null space of A as the span of a set of vectors, applying Theorem SSNS [283] as economically as possible, without reference to the underlying homogeneous system of equations (in contrast to Example SSNS [285]).

$$A = \begin{bmatrix} 2 & 1 & 5 & 1 & 5 & 1 \\ 1 & 1 & 3 & 1 & 6 & -1 \\ -1 & 1 & -1 & 0 & 4 & -3 \\ -3 & 2 & -4 & -4 & -7 & 0 \\ 3 & -1 & 5 & 2 & 2 & 3 \end{bmatrix}$$

Theorem SSNS [283] creates vectors for the span by first row-reducing the matrix in question.

 \boxtimes

The row-reduced version of A is

We will mechanically follow the prescription of Theorem SSNS [283]. Here we go, in two big steps.

First, the non-pivot columns have indices $F = \{3, 5, 6\}$, so we will construct the n-r = 6-3 = 3 vectors with a pattern of zeros and ones corresponding to the indices in F. This is the realization of the first two lines of the three-case definition of the vectors \mathbf{z}_j , $1 \le j \le n-r$.

$$\mathbf{z}_{1} = \begin{bmatrix} 1\\ 0\\ 0 \end{bmatrix} \qquad \qquad \mathbf{z}_{2} = \begin{bmatrix} 0\\ 1\\ 0 \end{bmatrix} \qquad \qquad \mathbf{z}_{3} = \begin{bmatrix} 0\\ 0\\ 1 \end{bmatrix}$$

Each of these vectors arises due to the presence of a column that is not a pivot column. The remaining entries of each vector are the entries of the corresponding non-pivot column, negated, and distributed into the empty slots in order (these slots have indices in the set D

and correspond to pivot columns). This is the realization of the third line of the three-case definition of the vectors \mathbf{z}_j , $1 \leq j \leq n-r$.

$$\mathbf{z}_{1} = \begin{bmatrix} -2\\ -1\\ 1\\ 0\\ 0\\ 0 \end{bmatrix} \qquad \mathbf{z}_{2} = \begin{bmatrix} 1\\ -3\\ 0\\ -4\\ 1\\ 0 \end{bmatrix} \qquad \mathbf{z}_{3} = \begin{bmatrix} -2\\ 1\\ 0\\ 2\\ 0\\ 1 \end{bmatrix}$$

So, by Theorem SSNS [283], we have

$$\mathcal{N}(A) = \langle \{\mathbf{z}_1, \, \mathbf{z}_2, \, \mathbf{z}_3\} \rangle = \left\langle \left\{ \begin{bmatrix} -2\\ -1\\ 1\\ 0\\ 0\\ 0 \end{bmatrix}, \, \begin{bmatrix} 1\\ -3\\ 0\\ -4\\ 1\\ 0 \end{bmatrix}, \, \begin{bmatrix} -2\\ 1\\ 0\\ 2\\ 0\\ 1 \end{bmatrix} \right\} \right\rangle$$

We know that the null space of A is the solution set of the homogeneous system $\mathcal{LS}(A, \mathbf{0})$, but nowhere in this application of Theorem SSNS [283] have we found occasion to reference the variables or equations of this system. These details are all buried in the proof of Theorem SSNS [283].

More advanced computational devices will compute the null space of a matrix. See:

Computation NS.MMA [1569] Here's an example that will simultaneously exercise the span construction and Theorem SSNS [283], while also pointing the way to the next section.

Example SCAD Span of the columns of Archetype D Begin with the set of four vectors of size 3

$$T = \{\mathbf{w}_1, \, \mathbf{w}_2, \, \mathbf{w}_3, \, \mathbf{w}_4\} = \left\{ \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 1\\4\\1 \end{bmatrix}, \begin{bmatrix} 7\\-5\\4 \end{bmatrix}, \begin{bmatrix} -7\\-6\\-5 \end{bmatrix} \right\}$$

and consider the infinite set $W = \langle T \rangle$. The vectors of T have been chosen as the four columns of the coefficient matrix in Archetype D [1670]. Check that the vector

$$\mathbf{z}_2 = \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix}$$

is a solution to the homogeneous system $\mathcal{LS}(D, \mathbf{0})$ (it is the vector \mathbf{z}_2 provided by the description of the null space of the coefficient matrix D from Theorem SSNS [283]). Applying Theorem SLSLC [229], we can write the linear combination,

$$2\mathbf{w}_1 + 3\mathbf{w}_2 + 0\mathbf{w}_3 + 1\mathbf{w}_4 = \mathbf{0}$$

which we can solve for \mathbf{w}_4 ,

$$\mathbf{w}_4 = (-2)\mathbf{w}_1 + (-3)\mathbf{w}_2.$$

This equation says that whenever we encounter the vector \mathbf{w}_4 , we can replace it with a specific linear combination of the vectors \mathbf{w}_1 and \mathbf{w}_2 . So using \mathbf{w}_4 in the set T, along with \mathbf{w}_1 and \mathbf{w}_2 , is excessive. An example of what we mean here can be illustrated by the computation,

$$5\mathbf{w}_{1} + (-4)\mathbf{w}_{2} + 6\mathbf{w}_{3} + (-3)\mathbf{w}_{4} = 5\mathbf{w}_{1} + (-4)\mathbf{w}_{2} + 6\mathbf{w}_{3} + (-3)((-2)\mathbf{w}_{1} + (-3)\mathbf{w}_{2})$$

= $5\mathbf{w}_{1} + (-4)\mathbf{w}_{2} + 6\mathbf{w}_{3} + (6\mathbf{w}_{1} + 9\mathbf{w}_{2})$
= $11\mathbf{w}_{1} + 5\mathbf{w}_{2} + 6\mathbf{w}_{3}$

So what began as a linear combination of the vectors \mathbf{w}_1 , \mathbf{w}_2 , \mathbf{w}_3 , \mathbf{w}_4 has been reduced to a linear combination of the vectors \mathbf{w}_1 , \mathbf{w}_2 , \mathbf{w}_3 . A careful proof using our definition of set equality (Definition SE [1604]) would now allow us to conclude that this reduction is possible for any vector in W, so

$$W = \langle \{\mathbf{w}_1, \, \mathbf{w}_2, \, \mathbf{w}_3\} \rangle$$

So the span of our set of vectors, W, has not changed, but we have *described* it by the span of a set of *three* vectors, rather than *four*. Furthermore, we can achieve yet another, similar, reduction.

Check that the vector

 $\mathbf{z}_1 = \begin{bmatrix} -3\\ -1\\ 1\\ 0 \end{bmatrix}$

is a solution to the homogeneous system $\mathcal{LS}(D, \mathbf{0})$ (it is the vector \mathbf{z}_1 provided by the description of the null space of the coefficient matrix D from Theorem SSNS [283]). Applying Theorem SLSLC [229], we can write the linear combination,

$$(-3)\mathbf{w}_1 + (-1)\mathbf{w}_2 + 1\mathbf{w}_3 = \mathbf{0}$$

which we can solve for \mathbf{w}_3 ,

 $\mathbf{w}_3 = 3\mathbf{w}_1 + 1\mathbf{w}_2$

This equation says that whenever we encounter the vector \mathbf{w}_3 , we can replace it with a specific linear combination of the vectors \mathbf{w}_1 and \mathbf{w}_2 . So, as before, the vector \mathbf{w}_3 is not needed in the description of W, provided we have \mathbf{w}_1 and \mathbf{w}_2 available. In particular, a careful proof (such as is done in Example RSC5 [366]) would show that

$$W = \langle \{\mathbf{w}_1, \, \mathbf{w}_2\} \rangle$$

So W began life as the span of a set of four vectors, and we have now shown (utilizing solutions to a homogeneous system) that W can also be described as the span of a set of just two vectors. Convince yourself that we cannot go any further. In other words, it is not

possible to dismiss either \mathbf{w}_1 or \mathbf{w}_2 in a similar fashion and winnow the set down to just one vector.

What was it about the original set of four vectors that allowed us to declare certain vectors as surplus? And just which vectors were we able to dismiss? And why did we have to stop once we had two vectors remaining? The answers to these questions motivate "linear independence," our next section and next definition, and so are worth considering carefully *now*.

It is possible to have your computational device crank out the vector form of the solution set to a linear system of equations. See: Computation VFSS.MMA [1571]

Subsection READ Reading Questions

1. Let S be the set of three vectors below.

$$S = \left\{ \begin{bmatrix} 1\\2\\-1 \end{bmatrix}, \begin{bmatrix} 3\\-4\\2 \end{bmatrix}, \begin{bmatrix} 4\\-2\\1 \end{bmatrix} \right\}$$

Let $W = \langle S \rangle$ be the span of S. Is the vector $\begin{bmatrix} -1 \\ 8 \\ -4 \end{bmatrix}$ in W? Give an explanation of the reason for your answer.

- 2. Use S and W from the previous question. Is the vector $\begin{bmatrix} 6\\5\\-1 \end{bmatrix}$ in W? Give an explanation of the reason for your answer.
- 3. For the matrix A below, find a set S so that $\langle S \rangle = \mathcal{N}(A)$, where $\mathcal{N}(A)$ is the null space of A. (See Theorem SSNS [283].)

$$A = \begin{bmatrix} 1 & 3 & 1 & 9 \\ 2 & 1 & -3 & 8 \\ 1 & 1 & -1 & 5 \end{bmatrix}$$

Subsection EXC Exercises

C22 For each archetype that is a system of equations, consider the corresponding homogeneous system of equations. Write elements of the solution set to these homogeneous systems in vector form, as guaranteed by Theorem VFSLS [244]. Then write the null space of the coefficient matrix of each system as the span of a set of vectors, as described in Theorem SSNS [283].

```
Archetype A [1642]
Archetype B [1652]
Archetype C [1662]
Archetype D [1670]/ Archetype E [1678]
Archetype F [1686]
Archetype G [1697]/ Archetype H [1705]
Archetype I [1714]
Archetype J [1723]
```

Contributed by Robert Beezer Solution [301]

C23 Archetype K [1734] and Archetype L [1742] are defined as matrices. Use Theorem SSNS [283] directly to find a set S so that $\langle S \rangle$ is the null space of the matrix. Do not make any reference to the associated homogeneous system of equations in your solution.

Contributed by Robert Beezer Solution [301]

C40 Suppose that
$$S = \left\{ \begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix}, \begin{bmatrix} 3\\2\\-2\\1 \end{bmatrix} \right\}$$
. Let $W = \langle S \rangle$ and let $\mathbf{x} = \begin{bmatrix} 5\\8\\-12\\-5 \end{bmatrix}$. Is $\mathbf{x} \in W$?

If so, provide an explicit linear combination that demonstrates this. Contributed by Robert Beezer Solution [301]

C41 Suppose that
$$S = \left\{ \begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix}, \begin{bmatrix} 3\\2\\-2\\1 \end{bmatrix} \right\}$$
. Let $W = \langle S \rangle$ and let $\mathbf{y} = \begin{bmatrix} 5\\1\\3\\5 \end{bmatrix}$. Is $\mathbf{y} \in W$? If so, provide an explicit linear combination that demonstrates this.

Contributed by Robert Beezer Solution [303]

C42 Suppose
$$R = \left\{ \begin{bmatrix} 2\\-1\\3\\4\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\2\\2\\-1\end{bmatrix}, \begin{bmatrix} 3\\-1\\0\\3\\-2 \end{bmatrix} \right\}$$
. Is $\mathbf{y} = \begin{bmatrix} 1\\-1\\-8\\-4\\-3 \end{bmatrix}$ in $\langle R \rangle$?
Contributed by Robert Beezer Solution [304]

C43 Suppose
$$R = \begin{cases} \begin{bmatrix} 2\\-1\\3\\4\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\2\\2\\-1 \end{bmatrix}, \begin{bmatrix} 3\\-1\\0\\3\\-2 \end{bmatrix} \end{cases}$$
. Is $\mathbf{z} = \begin{bmatrix} 1\\1\\5\\3\\1 \end{bmatrix}$ in $\langle R \rangle$?
Contributed by Robert Beezer Solution [305]
C44 Suppose that $S = \begin{cases} \begin{bmatrix} -1\\2\\1 \end{bmatrix}, \begin{bmatrix} 3\\1\\2\\1 \end{bmatrix}, \begin{bmatrix} 1\\5\\4\\2 \end{bmatrix}, \begin{bmatrix} -6\\5\\1\\2 \end{bmatrix} \end{cases}$. Let $W = \langle S \rangle$ and let $\mathbf{y} = \begin{bmatrix} -5\\3\\0 \end{bmatrix}$.
Is $\mathbf{y} \in W$? If so, provide an explicit linear combination that demonstrates this.
Contributed by Robert Beezer Solution [306]
C45 Suppose that $G = \begin{cases} \begin{bmatrix} -1\\2\\1 \end{bmatrix}, \begin{bmatrix} 3\\1\\2 \end{bmatrix}, \begin{bmatrix} 1\\5\\4 \end{bmatrix}, \begin{bmatrix} -6\\5\\1 \end{bmatrix} \end{cases}$. Let $W = \langle S \rangle$ and let $\mathbf{y} = \begin{bmatrix} -5\\3\\0 \end{bmatrix}$.

C45 Suppose that $S = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \end{bmatrix} \right\}$. Let $W = \langle S \rangle$ and let $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Is $\mathbf{w} \in W$? If so, provide an explicit linear combination that demonstrates this. Contributed by Robert Beezer Solution [307]

C50 Let A be the matrix below.

(a) Find a set S so that
$$\mathcal{N}(A) = \langle S \rangle$$
.

(b) If
$$\mathbf{z} = \begin{bmatrix} 3 \\ -5 \\ 1 \\ 2 \end{bmatrix}$$
, then show directly that $\mathbf{z} \in \mathcal{N}(A)$.

(c) Write \mathbf{z} as a linear combination of the vectors in S.

$$A = \begin{bmatrix} 2 & 3 & 1 & 4 \\ 1 & 2 & 1 & 3 \\ -1 & 0 & 1 & 1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [308]

C60 For the matrix A below, find a set of vectors S so that the span of S equals the null space of A, $\langle S \rangle = \mathcal{N}(A)$.

$$A = \begin{bmatrix} 1 & 1 & 6 & -8 \\ 1 & -2 & 0 & 1 \\ -2 & 1 & -6 & 7 \end{bmatrix}$$

Contributed by Robert Beezer Solution [310]

M10 Consider the set of all size 2 vectors in the Cartesian plane \mathbb{R}^2 .

- 1. Give a geometric description of the span of a single vector.
- 2. How can you tell if two vectors span the entire plane, without doing any row reduction or calculation?

Contributed by Chris Black Solution [311]

M11 Consider the set of all size 3 vectors in Cartesian 3-space \mathbb{R}^3 .

- 1. Give a geometric description of the span of a single vector.
- 2. Describe the possibilities for the span of two vectors.
- 3. Describe the possibilities for the span of three vectors.

Contributed by Chris Black Solution [312]

M12 Let
$$\mathbf{u} = \begin{bmatrix} 1\\ 3\\ -2 \end{bmatrix}$$
 and $\mathbf{v} = \begin{bmatrix} 2\\ -2\\ 1 \end{bmatrix}$.

- 1. Find a vector \mathbf{w}_1 , different from \mathbf{u} and \mathbf{v} , so that $\langle \{\mathbf{u}, \mathbf{v}, \mathbf{w}_1\} \rangle = \langle \{\mathbf{u}, \mathbf{v}\} \rangle$.
- 2. Find a vector \mathbf{w}_2 so that $\langle \{\mathbf{u}, \mathbf{v}, \mathbf{w}_2\} \rangle \neq \langle \{\mathbf{u}, \mathbf{v}\} \rangle$.

Contributed by Chris Black Solution [312]

M20 In Example SCAD [290] we began with the four columns of the coefficient matrix of Archetype D [1670], and used these columns in a span construction. Then we methodically argued that we could remove the last column, then the third column, and create the same set by just doing a span construction with the first two columns. We claimed we could not go any further, and had removed as many vectors as possible. Provide a convincing argument

for why a third vector cannot be removed. Contributed by Robert Beezer

M21 In the spirit of Example SCAD [290], begin with the four columns of the coefficient matrix of Archetype C [1662], and use these columns in a span construction to build the set S. Argue that S can be expressed as the span of just three of the columns of the coefficient matrix (saying exactly which three) and in the spirit of Exercise SS.M20 [299] argue that no one of these three vectors can be removed and still have a span construction create S. Contributed by Robert Beezer Solution [313]

T10 Suppose that $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{C}^m$. Prove that

$$\langle \{\mathbf{v}_1, \, \mathbf{v}_2\} \rangle = \langle \{\mathbf{v}_1, \, \mathbf{v}_2, \, 5\mathbf{v}_1 + 3\mathbf{v}_2\} \rangle$$

Contributed by Robert Beezer Solution [314]

T20 Suppose that S is a set of vectors from \mathbb{C}^m . Prove that the zero vector, **0**, is an element of $\langle S \rangle$.

Contributed by Robert Beezer Solution [315]

T21 Suppose that S is a set of vectors from \mathbb{C}^m and $\mathbf{x}, \mathbf{y} \in \langle S \rangle$. Prove that $\mathbf{x} + \mathbf{y} \in \langle S \rangle$. Contributed by Robert Beezer

T22 Suppose that S is a set of vectors from \mathbb{C}^m , $\alpha \in \mathbb{C}$, and $\mathbf{x} \in \langle S \rangle$. Prove that $\alpha \mathbf{x} \in \langle S \rangle$. Contributed by Robert Beezer

Subsection SOL Solutions

C22 Contributed by Robert Beezer Statement [295]

The vector form of the solutions obtained in this manner will involve precisely the vectors described in Theorem SSNS [283] as providing the null space of the coefficient matrix of the system as a span. These vectors occur in each archetype in a description of the null space. Studying Example VFSAL [252] may be of some help.

C23 Contributed by Robert Beezer Statement [295] Study Example NSDS [287] to understand the correct approach to this question. The solution for each is listed in the Archetypes (Appendix A [1637]) themselves.

C40 Contributed by Robert Beezer Statement [296] Rephrasing the question, we want to know if there are scalars α_1 and α_2 such that

$$\alpha_1 \begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3\\2\\-2\\1 \end{bmatrix} = \begin{bmatrix} 5\\8\\-12\\-5 \end{bmatrix}$$

Theorem SLSLC [229] allows us to rephrase the question again as a quest for solutions to

the system of four equations in two unknowns with an augmented matrix given by

$$\begin{bmatrix} 2 & 3 & 5 \\ -1 & 2 & 8 \\ 3 & -2 & -12 \\ 4 & 1 & -5 \end{bmatrix}$$

This matrix row-reduces to

$$\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

From the form of this matrix, we can see that $\alpha_1 = -2$ and $\alpha_2 = 3$ is an affirmative answer to our question. More convincingly,

$$(-2)\begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix} + (3)\begin{bmatrix} 3\\2\\-2\\1 \end{bmatrix} = \begin{bmatrix} 5\\8\\-12\\-5 \end{bmatrix}$$

C41 Contributed by Robert Beezer Statement [296]

Rephrasing the question, we want to know if there are scalars α_1 and α_2 such that

$$\alpha_1 \begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3\\2\\-2\\1 \end{bmatrix} = \begin{bmatrix} 5\\1\\3\\5 \end{bmatrix}$$

Theorem SLSLC [229] allows us to rephrase the question again as a quest for solutions to the system of four equations in two unknowns with an augmented matrix given by

$$\begin{bmatrix} 2 & 3 & 5 \\ -1 & 2 & 1 \\ 3 & -2 & 3 \\ 4 & 1 & 5 \end{bmatrix}$$

This matrix row-reduces to

$\left[1 \right]$	0	0]
0	1	0
0	0	1
0	0	0

With a leading 1 in the last column of this matrix (Theorem RCLS [120]) we can see that the system of equations has no solution, so there are no values for α_1 and α_2 that will allow us to conclude that **y** is in W. So $\mathbf{y} \notin W$. C42 Contributed by Robert Beezer Statement [296] Form a linear combination, with unknown scalars, of R that equals \mathbf{y} ,

$$a_{1}\begin{bmatrix}2\\-1\\3\\4\\0\end{bmatrix}+a_{2}\begin{bmatrix}1\\1\\2\\2\\-1\end{bmatrix}+a_{3}\begin{bmatrix}3\\-1\\0\\3\\-2\end{bmatrix}=\begin{bmatrix}1\\-1\\-8\\-4\\-3\end{bmatrix}$$

We want to know if there are values for the scalars that make the vector equation true since that is the definition of membership in $\langle R \rangle$. By Theorem SLSLC [229] any such values will also be solutions to the linear system represented by the augmented matrix,

$$\begin{bmatrix} 2 & 1 & 3 & 1 \\ -1 & 1 & -1 & -1 \\ 3 & 2 & 0 & -8 \\ 4 & 2 & 3 & -4 \\ 0 & -1 & -2 & -3 \end{bmatrix}$$

Row-reducing the matrix yields,

$$\begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From this we see that the system of equations is consistent (Theorem RCLS [120]), and has a unique solution. This solution will provide a linear combination of the vectors in R that equals \mathbf{y} . So $\mathbf{y} \in R$.

C43 Contributed by Robert Beezer Statement [297] Form a linear combination, with unknown scalars, of R that equals \mathbf{z} ,

$$a_{1}\begin{bmatrix}2\\-1\\3\\4\\0\end{bmatrix} + a_{2}\begin{bmatrix}1\\1\\2\\2\\-1\end{bmatrix} + a_{3}\begin{bmatrix}3\\-1\\0\\3\\-2\end{bmatrix} = \begin{bmatrix}1\\1\\5\\3\\1\end{bmatrix}$$

We want to know if there are values for the scalars that make the vector equation true since that is the definition of membership in $\langle R \rangle$. By Theorem SLSLC [229] any such values will also be solutions to the linear system represented by the augmented matrix,

$$\begin{bmatrix} 2 & 1 & 3 & 1 \\ -1 & 1 & -1 & 1 \\ 3 & 2 & 0 & 5 \\ 4 & 2 & 3 & 3 \\ 0 & -1 & -2 & 1 \end{bmatrix}$$

Row-reducing the matrix yields,

$$\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0
\end{bmatrix}$$

With a leading 1 in the last column, the system is inconsistent (Theorem RCLS [120]), so there are no scalars a_1 , a_2 , a_3 that will create a linear combination of the vectors in R that equal \mathbf{z} . So $\mathbf{z} \notin R$.

C44 Contributed by Robert Beezer Statement [297] Form a linear combination, with unknown scalars, of S that equals \mathbf{y} ,

$$a_1 \begin{bmatrix} -1\\2\\1 \end{bmatrix} + a_2 \begin{bmatrix} 3\\1\\2 \end{bmatrix} + a_3 \begin{bmatrix} 1\\5\\4 \end{bmatrix} + a_4 \begin{bmatrix} -6\\5\\1 \end{bmatrix} = \begin{bmatrix} -5\\3\\0 \end{bmatrix}$$

We want to know if there are values for the scalars that make the vector equation true since that is the definition of membership in $\langle S \rangle$. By Theorem SLSLC [229] any such values will also be solutions to the linear system represented by the augmented matrix,

$$\begin{bmatrix} -1 & 3 & 1 & -6 & -5 \\ 2 & 1 & 5 & 5 & 3 \\ 1 & 2 & 4 & 1 & 0 \end{bmatrix}$$

Row-reducing the matrix yields,

$$\begin{bmatrix} 1 & 0 & 2 & 3 & 2 \\ 0 & 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

From this we see that the system of equations is consistent (Theorem RCLS [120]), and has a infinitely many solutions. Any solution will provide a linear combination of the vectors in R that equals y. So $\mathbf{y} \in S$, for example,

$$(-10)\begin{bmatrix}-1\\2\\1\end{bmatrix} + (-2)\begin{bmatrix}3\\1\\2\end{bmatrix} + (3)\begin{bmatrix}1\\5\\4\end{bmatrix} + (2)\begin{bmatrix}-6\\5\\1\end{bmatrix} = \begin{bmatrix}-5\\3\\0\end{bmatrix}$$

C45 Contributed by Robert Beezer Statement [297] Form a linear combination, with unknown scalars, of S that equals \mathbf{w} ,

$$a_1 \begin{bmatrix} -1\\2\\1 \end{bmatrix} + a_2 \begin{bmatrix} 3\\1\\2 \end{bmatrix} + a_3 \begin{bmatrix} 1\\5\\4 \end{bmatrix} + a_4 \begin{bmatrix} -6\\5\\1 \end{bmatrix} = \begin{bmatrix} 2\\1\\3 \end{bmatrix}$$

We want to know if there are values for the scalars that make the vector equation true since that is the definition of membership in $\langle S \rangle$. By Theorem SLSLC [229] any such values will also be solutions to the linear system represented by the augmented matrix,

$$\begin{bmatrix} -1 & 3 & 1 & -6 & 2 \\ 2 & 1 & 5 & 5 & 1 \\ 1 & 2 & 4 & 1 & 3 \end{bmatrix}$$

Row-reducing the matrix yields,

1	0	2	3	0]
0		1	-1	$\begin{array}{c} 0 \\ 1 \end{array}$
0	0	0	0	1

With a leading 1 in the last column, the system is inconsistent (Theorem RCLS [120]), so there are no scalars a_1 , a_2 , a_3 , a_4 that will create a linear combination of the vectors in S that equal **w**. So $\mathbf{w} \notin \langle S \rangle$.

C50 Contributed by Robert Beezer Statement [297]

(a) Theorem SSNS [283] provides formulas for a set S with this property, but first we must row-reduce A

$$A \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

 x_3 and x_4 would be the free variables in the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ and Theorem

SSNS [283] provides the set $S = {\mathbf{z}_1, \mathbf{z}_2}$ where

$$\mathbf{z}_1 = \begin{bmatrix} 1\\ -1\\ 1\\ 0 \end{bmatrix} \qquad \qquad \mathbf{z}_2 = \begin{bmatrix} 1\\ -2\\ 0\\ 1 \end{bmatrix}$$

(b) Simply employ the components of the vector \mathbf{z} as the variables in the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. The three equations of this system evaluate as follows,

$$2(3) + 3(-5) + 1(1) + 4(2) = 0$$

$$1(3) + 2(-5) + 1(1) + 3(2) = 0$$

$$-1(3) + 0(-5) + 1(1) + 1(2) = 0$$

Since each result is zero, \mathbf{z} qualifies for membership in $\mathcal{N}(A)$.

(c) By Theorem SSNS [283] we know this must be possible (that is the moral of this exercise). Find scalars α_1 and α_2 so that

$$\alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2 = \alpha_1 \begin{bmatrix} 1\\ -1\\ 1\\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1\\ -2\\ 0\\ 1 \end{bmatrix} = \begin{bmatrix} 3\\ -5\\ 1\\ 2 \end{bmatrix} = \mathbf{z}$$

Theorem SLSLC [229] allows us to convert this question into a question about a system of

four equations in two variables. The augmented matrix of this system row-reduces to

A solution is $\alpha_1 = 1$ and $\alpha_2 = 2$. (Notice too that this solution is unique!)

C60 Contributed by Robert Beezer Statement [298] Theorem SSNS [283] says that if we find the vector form of the solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$, then the fixed vectors (one per free variable) will have the desired property. Row-reduce A, viewing it as the augmented matrix of a homogeneous system with an invisible columns of zeros as the last column,

$$\begin{bmatrix} 1 & 0 & 4 & -5 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Moving to the vector form of the solutions (Theorem VFSLS [244]), with free variables x_3 and x_4 , solutions to the consistent system (it is homogeneous, Theorem HSC [147]) can be expressed as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -4 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 5 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Then with S given by

$$S = \left\{ \begin{bmatrix} -4\\-2\\1\\0 \end{bmatrix}, \begin{bmatrix} 5\\3\\0\\1 \end{bmatrix} \right\}$$

Theorem SSNS [283] guarantees that

$$\mathcal{N}(A) = \langle S \rangle = \left\langle \left\{ \begin{bmatrix} -4\\-2\\1\\0 \end{bmatrix}, \begin{bmatrix} 5\\3\\0\\1 \end{bmatrix} \right\} \right\rangle$$

M10 Contributed by Chris Black Statement [298]

- 1. The span of a single vector \mathbf{v} is the set of all linear combinations of that vector. Thus, $\langle \mathbf{v} \rangle = \{ \alpha \mathbf{v} | \alpha \in \mathbb{R} \}$. This is the line through the origin and containing the (geometric) vector \mathbf{v} . Thus, if $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, then the span of \mathbf{v} is the line through (0,0) and (v_1, v_2) .
- 2. Two vectors will span the entire plane if they point in different directions, meaning that **u** does not lie on the line through **v** and vice-versa. That is, for vectors **u** and **v** in \mathbb{R}^2 , $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbb{R}^2$ if **u** is not a multiple of **v**.

M11 Contributed by Chris Black Statement [298]

- 1. The span of a single vector \mathbf{v} is the set of all linear combinations of that vector. Thus, $\langle \mathbf{v} \rangle = \{ \alpha \mathbf{v} | \alpha \in \mathbb{R} \}$. This is the line through the origin and containing the (geometric) vector \mathbf{v} . Thus, if $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$, then the span of \mathbf{v} is the line through (0, 0, 0) and (v_1, v_2, v_3) .
- 2. If the two vectors point in the same direction, then their span is the line through them. Recall that while two points determine a line, three points determine a plane. Two vectors will span a plane if they point in different directions, meaning that \mathbf{u} does not lie on the line through \mathbf{v} and vice-versa. The plane spanned by $\mathbf{u} = \begin{bmatrix} u_1 \\ u_1 \\ u_1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ is determined by the origin and the points (u_1, u_2, u_3) and (v_1, v_2, v_3) .
- 3. If all three vectors lie on the same line, then the span is that line. If one is a linear combination of the other two, but they are not all on the same line, then they will lie in a plane. Otherwise, the span of the set of three vectors will be all of 3-space.
- M12 Contributed by Chris Black Statement [299]

- 1. If we can find a vector \mathbf{w}_1 that is a linear combination of \mathbf{u} and \mathbf{v} , then $\langle \{\mathbf{u}, \mathbf{v}, \mathbf{w}_1\} \rangle$ will be the same set as $\langle \{\mathbf{u}, \mathbf{v}\} \rangle$. Thus, \mathbf{w}_1 can be any linear combination of \mathbf{u} and \mathbf{v} . One such example is $\mathbf{w}_1 = 3\mathbf{u} - \mathbf{v} = \begin{bmatrix} 1\\11\\-7 \end{bmatrix}$.
- 2. Now we are looking for a vector \mathbf{w}_2 that cannot be written as a linear combination of \mathbf{u} and \mathbf{v} . How can we find such a vector? Any vector that matches two components but not the third of any element of $\langle \{\mathbf{u}, \mathbf{v}\} \rangle$ will not be in the span. (Why?) One such example is $\mathbf{w}_2 = \begin{bmatrix} 4 \\ -4 \\ 1 \end{bmatrix}$ (which is nearly $2\mathbf{v}$, but not quite).

M21 Contributed by Robert Beezer Statement [300] If the columns of the coefficient matrix from Archetype C [1662] are named \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 , \mathbf{u}_4 then we can discover the equation

$$(-2)\mathbf{u}_1 + (-3)\mathbf{u}_2 + \mathbf{u}_3 + \mathbf{u}_4 = \mathbf{0}$$

by building a homogeneous system of equations and viewing a solution to the system as scalars in a linear combination via Theorem SLSLC [229]. This particular vector equation can be rearranged to read

$$\mathbf{u}_4 = (2)\mathbf{u}_1 + (3)\mathbf{u}_2 + (-1)\mathbf{u}_3$$

This can be interpreted to mean that \mathbf{u}_4 is unnecessary in $\langle \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\} \rangle$, so that

$$\langle \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3, \, \mathbf{u}_4 \} \rangle = \langle \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3 \} \rangle$$

If we try to repeat this process and find a linear combination of \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 that equals the zero vector, we will fail. The required homogeneous system of equations (via Theorem SLSLC [229]) has only a trivial solution, which will not provide the kind of equation we need to remove one of the three remaining vectors.

T10 Contributed by Robert Beezer Statement [300] This is an equality of sets, so Definition SE [1604] applies.

First show that $X = \langle \{\mathbf{v}_1, \mathbf{v}_2\} \rangle \subseteq \langle \{\mathbf{v}_1, \mathbf{v}_2, 5\mathbf{v}_1 + 3\mathbf{v}_2\} \rangle = Y$. Choose $\mathbf{x} \in X$. Then $\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$ for some scalars a_1 and a_2 . Then,

$$\mathbf{x} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + 0(5\mathbf{v}_1 + 3\mathbf{v}_2)$$

which qualifies **x** for membership in Y, as it is a linear combination of \mathbf{v}_1 , \mathbf{v}_2 , $5\mathbf{v}_1 + 3\mathbf{v}_2$.

Now show the opposite inclusion, $Y = \langle \{\mathbf{v}_1, \mathbf{v}_2, 5\mathbf{v}_1 + 3\mathbf{v}_2\} \rangle \subseteq \langle \{\mathbf{v}_1, \mathbf{v}_2\} \rangle = X$. Choose $\mathbf{y} \in Y$. Then there are scalars a_1, a_2, a_3 such that

$$\mathbf{y} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 (5 \mathbf{v}_1 + 3 \mathbf{v}_2)$$

Rearranging, we obtain,

$$\mathbf{y} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 (5 \mathbf{v}_1 + 3 \mathbf{v}_2)$$

 $= a_{1}\mathbf{v}_{1} + a_{2}\mathbf{v}_{2} + 5a_{3}\mathbf{v}_{1} + 3a_{3}\mathbf{v}_{2}$ Property DVAC [206] $= a_{1}\mathbf{v}_{1} + 5a_{3}\mathbf{v}_{1} + a_{2}\mathbf{v}_{2} + 3a_{3}\mathbf{v}_{2}$ Property CC [205] $= (a_{1} + 5a_{3})\mathbf{v}_{1} + (a_{2} + 3a_{3})\mathbf{v}_{2}$ Property DSAC [206]

This is an expression for \mathbf{y} as a linear combination of \mathbf{v}_1 and \mathbf{v}_2 , earning \mathbf{y} membership in X. Since X is a subset of Y, and vice versa, we see that X = Y, as desired.

T20 Contributed by Robert Beezer Statement [300] No matter what the elements of the set S are, we can choose the scalars in a linear combination to all be zero. Suppose that $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_p}$. Then compute

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + 0\mathbf{v}_3 + \dots + 0\mathbf{v}_p = \mathbf{0} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0}$$
$$= \mathbf{0}$$

But what if we choose S to be the empty set? The *convention* is that the empty sum in Definition SSCV [270] evaluates to "zero," in this case this is the zero vector.

Section LI Linear Independence

Subsection LISV Linearly Independent Sets of Vectors

Theorem SLSLC [229] tells us that a solution to a homogeneous system of equations is a linear combination of the columns of the coefficient matrix that equals the zero vector. We used just this situation to our advantage (twice!) in Example SCAD [290] where we reduced the set of vectors used in a span construction from four down to two, by declaring certain vectors as surplus. The next two definitions will allow us to formalize this situation.

Definition RLDCV Relation of Linear Dependence for Column Vectors Given a set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$, a true statement of the form

 $\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \dots + \alpha_n\mathbf{u}_n = \mathbf{0}$

is a relation of linear dependence on S. If this statement is formed in a trivial fashion, i.e. $\alpha_i = 0, 1 \le i \le n$, then we say it is the trivial relation of linear dependence on S.

\triangle

Definition LICV Linear Independence of Column Vectors

The set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is **linearly dependent** if there is a relation of linear dependence on S that is not trivial. In the case where the *only* relation of linear dependence on S is the trivial one, then S is a **linearly independent** set of vectors. \triangle

Notice that a relation of linear dependence is an *equation*. Though most of it is a linear combination, it is not a linear combination (that would be a vector). Linear independence is a property of a *set* of vectors. It is easy to take a set of vectors, and an equal number of scalars, *all zero*, and form a linear combination that equals the zero vector. When the easy way is the *only* way, then we say the set is linearly independent. Here's a couple of examples.

Example LDS

Linearly dependent set in \mathbb{C}^5

Consider the set of n = 4 vectors from \mathbb{C}^5 ,

$$S = \left\{ \begin{bmatrix} 2\\-1\\3\\1\\2 \end{bmatrix}, \begin{bmatrix} 1\\2\\-1\\5\\2 \end{bmatrix}, \begin{bmatrix} 2\\1\\-3\\6\\1 \end{bmatrix}, \begin{bmatrix} -6\\7\\-1\\0\\1 \end{bmatrix} \right\}$$

To determine linear independence we first form a relation of linear dependence,

$$\alpha_{1} \begin{bmatrix} 2\\-1\\3\\1\\2 \end{bmatrix} + \alpha_{2} \begin{bmatrix} 1\\2\\-1\\5\\2 \end{bmatrix} + \alpha_{3} \begin{bmatrix} 2\\1\\-3\\6\\1 \end{bmatrix} + \alpha_{4} \begin{bmatrix} -6\\7\\-1\\0\\1 \end{bmatrix} = \mathbf{0}$$

We know that $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ is a solution to this equation, but that is of no interest whatsoever. That is *always* the case, no matter what four vectors we might have chosen. We are curious to know if there are other, nontrivial, solutions. Theorem SLSLC [229] tells us that we can find such solutions as solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ where the coefficient matrix has these four vectors as columns,

$$A = \begin{bmatrix} 2 & 1 & 2 & -6 \\ -1 & 2 & 1 & 7 \\ 3 & -1 & -3 & -1 \\ 1 & 5 & 6 & 0 \\ 2 & 2 & 1 & 1 \end{bmatrix}$$

Row-reducing this coefficient matrix yields,

$$\begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We could solve this homogeneous system completely, but for this example all we need is one nontrivial solution. Setting the lone free variable to any nonzero value, such as $x_4 = 1$, yields the nontrivial solution

$$\mathbf{x} = \begin{bmatrix} 2\\ -4\\ 3\\ 1 \end{bmatrix}$$

completing our application of Theorem SLSLC [229], we have

$$2\begin{bmatrix}2\\-1\\3\\1\\2\end{bmatrix} + (-4)\begin{bmatrix}1\\2\\-1\\5\\2\end{bmatrix} + 3\begin{bmatrix}2\\1\\-3\\6\\1\end{bmatrix} + 1\begin{bmatrix}-6\\7\\-1\\0\\1\end{bmatrix} = \mathbf{0}$$

This is a relation of linear dependence on S that is not trivial, so we conclude that S is

linearly dependent.

Example LIS Linearly independent set in \mathbb{C}^5 Consider the set of n = 4 vectors from \mathbb{C}^5 ,

$$T = \left\{ \begin{bmatrix} 2\\-1\\3\\1\\2 \end{bmatrix}, \begin{bmatrix} 1\\2\\-1\\5\\2 \end{bmatrix}, \begin{bmatrix} 2\\1\\-3\\6\\1 \end{bmatrix}, \begin{bmatrix} -6\\7\\-1\\1\\1 \end{bmatrix} \right\}$$

To determine linear independence we first form a relation of linear dependence,

$$\alpha_{1} \begin{bmatrix} 2\\-1\\3\\1\\2 \end{bmatrix} + \alpha_{2} \begin{bmatrix} 1\\2\\-1\\5\\2 \end{bmatrix} + \alpha_{3} \begin{bmatrix} 2\\1\\-3\\6\\1 \end{bmatrix} + \alpha_{4} \begin{bmatrix} -6\\7\\-1\\1\\1 \end{bmatrix} = \mathbf{0}$$

We know that $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ is a solution to this equation, but that is of no interest whatsoever. That is *always* the case, no matter what four vectors we might have chosen. We are curious to know if there are other, nontrivial, solutions. Theorem SLSLC [229] tells us that we can find such solutions as solution to the homogeneous system $\mathcal{LS}(B, \mathbf{0})$ where the

 \boxtimes

coefficient matrix has these four vectors as columns. Row-reducing this coefficient matrix yields,

$$B = \begin{bmatrix} 2 & 1 & 2 & -6 \\ -1 & 2 & 1 & 7 \\ 3 & -1 & -3 & -1 \\ 1 & 5 & 6 & 1 \\ 2 & 2 & 1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

From the form of this matrix, we see that there are no free variables, so the solution is unique, and because the system is homogeneous, this unique solution is the trivial solution. So we now know that there is but one way to combine the four vectors of T into a relation of linear dependence, and that one way is the easy and obvious way. In this situation we say that the set, T, is linearly independent.

Example LDS [317] and Example LIS [320] relied on solving a homogeneous system of equations to determine linear independence. We can codify this process in a time-saving theorem.

Theorem LIVHS

Linearly Independent Vectors and Homogeneous Systems

Suppose that A is an $m \times n$ matrix and $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_n\}$ is the set of vectors in \mathbb{C}^m that are the columns of A. Then S is a linearly independent set if and only if the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution.

Proof (\Leftarrow) Suppose that $\mathcal{LS}(A, \mathbf{0})$ has a unique solution. Since it is a homogeneous system,

this solution must be the trivial solution $\mathbf{x} = \mathbf{0}$. By Theorem SLSLC [229], this means that the only relation of linear dependence on S is the trivial one. So S is linearly independent.

 (\Rightarrow) We will prove the contrapositive. Suppose that $\mathcal{LS}(A, \mathbf{0})$ does not have a unique solution. Since it is a homogeneous system, it is consistent (Theorem HSC [147]), and so must have infinitely many solutions (Theorem PSSLS [126]). One of these infinitely many solutions must be nontrivial (in fact, almost all of them are), so choose one. By Theorem SLSLC [229] this nontrivial solution will give a nontrivial relation of linear dependence on S, so we can conclude that S is a linearly dependent set.

Since Theorem LIVHS [321] is an equivalence, we can use it to determine the linear independence or dependence of any set of column vectors, just by creating a corresponding matrix and analyzing the row-reduced form. Let's illustrate this with two more examples.

Example LIHS Linearly independent, homogeneous system Is the set of vectors

Is the set of vectors

$$S = \left\{ \begin{bmatrix} 2\\-1\\3\\4\\2 \end{bmatrix}, \begin{bmatrix} 6\\2\\-1\\3\\4 \end{bmatrix}, \begin{bmatrix} 4\\3\\-4\\5\\1 \end{bmatrix} \right\}$$

linearly independent or linearly dependent?

Theorem LIVHS [321] suggests we study the matrix whose columns are the vectors in S,

$$A = \begin{bmatrix} 2 & 6 & 4 \\ -1 & 2 & 3 \\ 3 & -1 & -4 \\ 4 & 3 & 5 \\ 2 & 4 & 1 \end{bmatrix}$$

Specifically, we are interested in the size of the solution set for the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. Row-reducing A, we obtain

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Now, r = 3, so there are n - r = 3 - 3 = 0 free variables and we see that $\mathcal{LS}(A, \mathbf{0})$ has a unique solution (Theorem HSC [147], Theorem FVCS [124]). By Theorem LIVHS [321], the set S is linearly independent.

Example LDHS Linearly dependent, homogeneous system Is the set of vectors

$$S = \left\{ \begin{bmatrix} 2\\-1\\3\\4\\2 \end{bmatrix}, \begin{bmatrix} 6\\2\\-1\\3\\4 \end{bmatrix}, \begin{bmatrix} 4\\3\\-4\\-1\\2 \end{bmatrix} \right\}$$

linearly independent or linearly dependent?

Theorem LIVHS [321] suggests we study the matrix whose columns are the vectors in S,

$$A = \begin{bmatrix} 2 & 6 & 4 \\ -1 & 2 & 3 \\ 3 & -1 & -4 \\ 4 & 3 & -1 \\ 2 & 4 & 2 \end{bmatrix}$$

Specifically, we are interested in the size of the solution set for the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. Row-reducing A, we obtain

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Now, r = 2, so there are n - r = 3 - 2 = 1 free variables and we see that $\mathcal{LS}(A, \mathbf{0})$ has infinitely many solutions (Theorem HSC [147], Theorem FVCS [124]). By Theorem LIVHS [321], the set S is linearly dependent.

As an equivalence, Theorem LIVHS [321] gives us a straightforward way to determine if a set of vectors is linearly independent or dependent.

Review Example LIHS [322] and Example LDHS [323]. They are very similar, differing only in the last two slots of the third vector. This resulted in slightly different matrices when row-reduced, and slightly different values of r, the number of nonzero rows. Notice, too, that we are less interested in the actual solution set, and more interested in its form or size. These observations allow us to make a slight improvement in Theorem LIVHS [321].

Theorem LIVRN

Linearly Independent Vectors, r and n

Suppose that A is an $m \times n$ matrix and $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_n\}$ is the set of vectors in \mathbb{C}^m that are the columns of A. Let B be a matrix in reduced row-echelon form that is row-equivalent to A and let r denote the number of non-zero rows in B. Then S is linearly independent if and only if n = r.

Proof Theorem LIVHS [321] says the linear independence of S is equivalent to the homogeneous linear system $\mathcal{LS}(A, \mathbf{0})$ having a unique solution. Since $\mathcal{LS}(A, \mathbf{0})$ is consistent (Theorem HSC [147]) we can apply Theorem CSRN [123] to see that the solution is unique exactly when n = r.

So now here's an example of the most straightforward way to determine if a set of column vectors in linearly independent or linearly dependent. While this method can be quick and

easy, don't forget the logical progression from the definition of linear independence through homogeneous system of equations which makes it possible.

Example LDRN Linearly dependent, r < nIs the set of vectors

$$S = \left\{ \begin{bmatrix} 2\\-1\\3\\1\\0\\3 \end{bmatrix}, \begin{bmatrix} 9\\-6\\-2\\1\\2\\3\\2\\1 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\0\\0\\1\\1 \end{bmatrix}, \begin{bmatrix} -3\\1\\4\\2\\1\\2\\1\\2 \end{bmatrix}, \begin{bmatrix} 6\\-2\\1\\4\\3\\2\\2 \end{bmatrix} \right\}$$

linearly independent or linearly dependent? Theorem LIVHS [321] suggests we place these vectors into a matrix as columns and analyze the row-reduced version of the matrix,

2	9	1	-3	6]		$\lceil 1 \rceil$	0	0	0	-1
-1	-6	1	1	-2		0		0	0	1
3	-2	1	4	1	RREF	0	0	1	0	2
1	3	0	2	4	\rightarrow	0	0	0	1	1
0	2	0	1	3		0	0	0	0	
3	1	1	2	2		0	0	0	0	0

Now we need only compute that r = 4 < 5 = n to recognize, via Theorem LIVHS [321] that

S is a linearly dependent set. Boom!

Example LLDS Large linearly dependent set in \mathbb{C}^4 Consider the set of n = 9 vectors from \mathbb{C}^4 ,

$$R = \left\{ \begin{bmatrix} -1\\3\\1\\2 \end{bmatrix}, \begin{bmatrix} 7\\1\\-3\\6 \end{bmatrix}, \begin{bmatrix} 1\\2\\-1\\-2 \end{bmatrix}, \begin{bmatrix} 0\\4\\2\\9 \end{bmatrix}, \begin{bmatrix} 5\\-2\\4\\3 \end{bmatrix}, \begin{bmatrix} 2\\1\\-6\\4 \end{bmatrix}, \begin{bmatrix} 3\\0\\-3\\1 \end{bmatrix}, \begin{bmatrix} 1\\1\\5\\3 \end{bmatrix}, \begin{bmatrix} -6\\-1\\1\\1 \end{bmatrix} \right\}.$$

To employ Theorem LIVHS [321], we form a 4×9 coefficient matrix, C,

$$C = \begin{bmatrix} -1 & 7 & 1 & 0 & 5 & 2 & 3 & 1 & -6 \\ 3 & 1 & 2 & 4 & -2 & 1 & 0 & 1 & -1 \\ 1 & -3 & -1 & 2 & 4 & -6 & -3 & 5 & 1 \\ 2 & 6 & -2 & 9 & 3 & 4 & 1 & 3 & 1 \end{bmatrix}$$

To determine if the homogeneous system $\mathcal{LS}(C, \mathbf{0})$ has a unique solution or not, we would normally row-reduce this matrix. But in this particular example, we can do better. Theorem HMVEI [151] tells us that since the system is homogeneous with n = 9 variables in m = 4equations, and n > m, there must be infinitely many solutions. Since there is not a unique solution, Theorem LIVHS [321] says the set is linearly dependent.

The situation in Example LLDS [327] is slick enough to warrant formulating as a theorem.

 \boxtimes

Theorem MVSLD More Vectors than Size implies Linear Dependence

Suppose that $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is the set of vectors in \mathbb{C}^m , and that n > m. Then S is a linearly dependent set.

Proof Form the $m \times n$ coefficient matrix A that has the column vectors \mathbf{u}_i , $1 \leq i \leq n$ as its columns. Consider the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. By Theorem HMVEI [151] this system has infinitely many solutions. Since the system does not have a unique solution, Theorem LIVHS [321] says the columns of A form a linearly dependent set, which is the desired conclusion.

Subsection LINM Linear Independence and Nonsingular Matrices

We will now specialize to sets of n vectors from \mathbb{C}^n . This will put Theorem MVSLD [328] offlimits, while Theorem LIVHS [321] will involve square matrices. Let's begin by contrasting Archetype A [1642] and Archetype B [1652].

Example LDCAA Linearly dependent columns in Archetype A Archetype A [1642] is a system of linear equations with coefficient matrix,

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Do the columns of this matrix form a linearly independent or dependent set? By Example S [174] we know that A is singular. According to the definition of nonsingular matrices, Definition NM [173], the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has infinitely many solutions. So by Theorem LIVHS [321], the columns of A form a linearly dependent set.

Example LICAB Linearly independent columns in Archetype B

Archetype B [1652] is a system of linear equations with coefficient matrix,

$$B = \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix}$$

Do the columns of this matrix form a linearly independent or dependent set? By Example NM [174] we know that B is nonsingular. According to the definition of nonsingular matrices, Definition NM [173], the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution. So by Theorem LIVHS [321], the columns of B form a linearly independent set.

That Archetype A [1642] and Archetype B [1652] have opposite properties for the columns of their coefficient matrices is no accident. Here's the theorem, and then we will update our equivalences for nonsingular matrices, Theorem NME1 [181].

Theorem NMLIC

Nonsingular Matrices have Linearly Independent Columns

Suppose that A is a square matrix. Then A is nonsingular if and only if the columns of A form a linearly independent set. \Box

Proof This is a proof where we can chain together equivalences, rather than proving the two halves separately.

A nonsingular $\nleftrightarrow \mathcal{LS}(A, \mathbf{0})$ has a unique solutionDefinition NM [173] \iff columns of A are linearly independentTheorem LIVHS [321]

Here's an update to Theorem NME1 [181].

Theorem NME2

Nonsingular Matrix Equivalences, Round 2

Suppose that A is a square matrix. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .

5. The columns of A form a linearly independent set.

Proof Theorem NMLIC [330] is yet another equivalence for a nonsingular matrix, so we can add it to the list in Theorem NME1 [181].

Subsection NSSLI Null Spaces, Spans, Linear Independence

In Subsection SS.SSNS [283] we proved Theorem SSNS [283] which provided n - r vectors that could be used with the span construction to build the entire null space of a matrix. As we have hinted in Example SCAD [290], and as we will see again going forward, linearly dependent sets carry redundant vectors with them when used in building a set as a span. Our aim now is to show that the vectors provided by Theorem SSNS [283] form a linearly independent set, so in one sense they are as efficient as possible a way to describe the null space. Notice that the vectors \mathbf{z}_j , $1 \le j \le n - r$ first appear in the vector form of solutions to arbitrary linear systems (Theorem VFSLS [244]). The exact same vectors appear again in the span construction in the conclusion of Theorem SSNS [283]. Since this second theorem specializes to homogeneous systems the only real difference is that the vector \mathbf{c} in Theorem VFSLS [244] is the zero vector for a homogeneous system. Finally, Theorem BNS [334] will now show that these same vectors are a linearly independent set. We'll set the stage for the

 \square

proof of this theorem with a moderately large example. Study the example carefully, as it will make it easier to understand the proof.

Example LINSB

Linear independence of null space basis

Suppose that we are interested in the null space of a 3×7 matrix, A, which row-reduces to

$$B = \begin{bmatrix} 1 & 0 & -2 & 4 & 0 & 3 & 9 \\ 0 & 1 & 5 & 6 & 0 & 7 & 1 \\ 0 & 0 & 0 & 0 & 1 & 8 & -5 \end{bmatrix}$$

The set $F = \{3, 4, 6, 7\}$ is the set of indices for our four free variables that would be used in a description of the solution set for the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. Applying Theorem SSNS [283] we can begin to construct a set of four vectors whose span is the null space of A, a set of vectors we will reference as T.

$$\mathcal{N}(A) = \langle T \rangle = \langle \{ \mathbf{z}_1, \, \mathbf{z}_2, \, \mathbf{z}_3, \, \mathbf{z}_4 \} \rangle = \left\langle \left\{ \begin{bmatrix} 1\\0\\0\\0\\0\end{bmatrix}, \, \begin{bmatrix} 0\\0\\1\\0\\0\\0\end{bmatrix}, \, \begin{bmatrix} 0\\0\\1\\0\\0\\1\\0 \end{bmatrix}, \, \begin{bmatrix} 0\\0\\0\\1\\0\\1 \end{bmatrix} \right\} \right\rangle$$

So far, we have constructed as much of these individual vectors as we can, based just on the knowledge of the contents of the set F. This has allowed us to determine the entries in slots

3, 4, 6 and 7, while we have left slots 1, 2 and 5 blank. Without doing any more, lets ask if T is linearly independent? Begin with a relation of linear dependence on T, and see what we can learn about the scalars,

Applying Definition CVE [199] to the two ends of this chain of equalities, we see that $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$. So the only relation of linear dependence on the set T is a trivial one. By Definition LICV [317] the set T is linearly independent. The important feature of

this example is how the "pattern of zeros and ones" in the four vectors led to the conclusion of linear independence. $\hfill \boxtimes$

The proof of Theorem BNS [334] is really quite straightforward, and relies on the "pattern of zeros and ones" that arise in the vectors \mathbf{z}_i , $1 \leq i \leq n - r$ in the entries that correspond to the free variables. Play along with Example LINSB [332] as you study the proof. Also, take a look at Example VFSAD [232], Example VFSAI [248] and Example VFSAL [252], especially at the conclusion of Step 2 (temporarily ignore the construction of the constant vector, \mathbf{c}). This proof is also a good first example of how to prove a conclusion that states a set is linearly independent.

Theorem BNS Basis for Null Spaces

Suppose that A is an $m \times n$ matrix, and B is a row-equivalent matrix in reduced row-echelon form with r nonzero rows. Let $D = \{d_1, d_2, d_3, \ldots, d_r\}$ and $F = \{f_1, f_2, f_3, \ldots, f_{n-r}\}$ be the sets of column indices where B does and does not (respectively) have leading 1's. Construct the n - r vectors \mathbf{z}_j , $1 \le j \le n - r$ of size n as

$$\begin{bmatrix} \mathbf{z}_j \end{bmatrix}_i = \begin{cases} 1 & \text{if } i \in F, \ i = f_j \\ 0 & \text{if } i \in F, \ i \neq f_j \\ -\begin{bmatrix} B \end{bmatrix}_{k, f_j} & \text{if } i \in D, \ i = d_k \end{cases}$$

Define the set $S = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \dots, \mathbf{z}_{n-r}\}$. Then

1. $\mathcal{N}(A) = \langle S \rangle$.

2. S is a linearly independent set.

Proof Notice first that the vectors \mathbf{z}_j , $1 \leq j \leq n-r$ are exactly the same as the n-r vectors defined in Theorem SSNS [283]. Also, the hypotheses of Theorem SSNS [283] are the same as the hypotheses of the theorem we are currently proving. So it is then simply the conclusion of Theorem SSNS [283] that tells us that $\mathcal{N}(A) = \langle S \rangle$. That was the easy half, but the second part is not much harder. What is new here is the claim that S is a linearly independent set.

To prove the linear independence of a set, we need to start with a relation of linear dependence and somehow conclude that the scalars involved *must all be zero*, i.e. that the relation of linear dependence only happens in the trivial fashion. So to establish the linear independence of S, we start with

$$\alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2 + \alpha_3 \mathbf{z}_3 + \dots + \alpha_{n-r} \mathbf{z}_{n-r} = \mathbf{0}.$$

For each $j, 1 \leq j \leq n - r$, consider the equality of the individual entries of the vectors on both sides of this equality in position f_j ,

$$0 = [\mathbf{0}]_{f_j}$$

$$= [\alpha_1 \mathbf{z}_1 + \alpha_2 \mathbf{z}_2 + \alpha_3 \mathbf{z}_3 + \dots + \alpha_{n-r} \mathbf{z}_{n-r}]_{f_j}$$
Definition CVE [199]
$$= [\alpha_1 \mathbf{z}_1]_{f_j} + [\alpha_2 \mathbf{z}_2]_{f_j} + [\alpha_3 \mathbf{z}_3]_{f_j} + \dots + [\alpha_{n-r} \mathbf{z}_{n-r}]_{f_j}$$
Definition CVA [201]
$$= \alpha_1 [\mathbf{z}_1]_{f_j} + \alpha_2 [\mathbf{z}_2]_{f_j} + \alpha_3 [\mathbf{z}_3]_{f_j} + \dots +$$

 \square

$$\alpha_{j-1} [\mathbf{z}_{j-1}]_{f_j} + \alpha_j [\mathbf{z}_j]_{f_j} + \alpha_{j+1} [\mathbf{z}_{j+1}]_{f_j} + \dots + \alpha_{n-r} [\mathbf{z}_{n-r}]_{f_j} \qquad \text{Definition CVSM [203]}$$

$$= \alpha_1(0) + \alpha_2(0) + \alpha_3(0) + \dots + \alpha_{n-r}(0) \qquad \text{Definition of } \mathbf{z}_j$$

$$= \alpha_j$$

So for all $j, 1 \leq j \leq n-r$, we have $\alpha_j = 0$, which is the conclusion that tells us that the only relation of linear dependence on $S = \{\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \ldots, \mathbf{z}_{n-r}\}$ is the trivial one. Hence, by Definition LICV [317] the set is linearly independent, as desired.

Example NSLIL

Null space spanned by linearly independent set, Archetype L

In Example VFSAL [252] we previewed Theorem SSNS [283] by finding a set of two vectors such that their span was the null space for the matrix in Archetype L [1742]. Writing the matrix as L, we have

$$\mathcal{N}(L) = \left\langle \left\{ \begin{bmatrix} -1\\2\\-2\\1\\0 \end{bmatrix}, \begin{bmatrix} 2\\-2\\1\\0\\1 \end{bmatrix} \right\} \right\rangle$$

Solving the homogeneous system $\mathcal{LS}(L, \mathbf{0})$ resulted in recognizing x_4 and x_5 as the free variables. So look in entries 4 and 5 of the two vectors above and notice the pattern of zeros

and ones that provides the linear independence of the set.

Subsection READ Reading Questions

1. Let S be the set of three vectors below.

$$S = \left\{ \begin{bmatrix} 1\\2\\-1 \end{bmatrix}, \begin{bmatrix} 3\\-4\\2 \end{bmatrix}, \begin{bmatrix} 4\\-2\\1 \end{bmatrix} \right\}$$

Is S linearly independent or linearly dependent? Explain why.

2. Let S be the set of three vectors below.

$$S = \left\{ \begin{bmatrix} 1\\-1\\0 \end{bmatrix}, \begin{bmatrix} 3\\2\\2 \end{bmatrix}, \begin{bmatrix} 4\\3\\-4 \end{bmatrix} \right\}$$

Is S linearly independent or linearly dependent? Explain why.

 \boxtimes

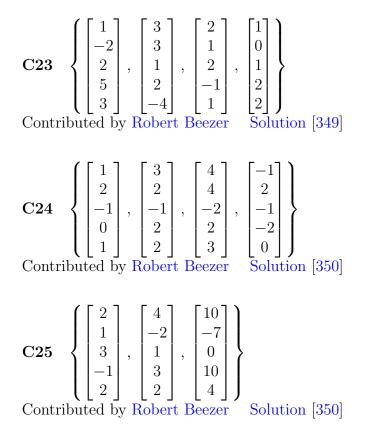
3. Based on your answer to the previous question, is the matrix below singular or nonsingular? Explain.

$$\begin{bmatrix} 1 & 3 & 4 \\ -1 & 2 & 3 \\ 0 & 2 & -4 \end{bmatrix}$$

Subsection EXC Exercises

Determine if the sets of vectors in Exercises C20–C25 are linearly independent or linearly dependent. When the set is linearly dependent, exhibit a nontrivial relation of linear dependence.

C20 $\left\{ \begin{vmatrix} 1 \\ -2 \\ 1 \end{vmatrix}, \begin{vmatrix} 2 \\ -1 \\ 3 \end{vmatrix}, \begin{vmatrix} 1 \\ 5 \\ 0 \end{vmatrix} \right\}$ Contributed by Robert Beezer Solution [348] $\mathbf{C21} \quad \left\{ \begin{array}{c|c} 1 \\ 2 \\ 4 \end{array} \right|, \begin{array}{c|c} 3 \\ -1 \\ -1 \end{array} \right|, \begin{array}{c|c} 1 \\ 3 \\ -6 \end{array} \right\}$ Contributed by Robert Beezer Solution [348] C22 $\left\{ \begin{array}{c} -2 \\ -1 \\ -1 \end{array} \right\}, \begin{array}{c} 1 \\ 0 \\ -1 \end{array} \right\}, \begin{array}{c} 3 \\ 3 \\ 6 \end{array} \right\}, \begin{array}{c} -5 \\ -4 \\ -6 \end{array} \right\}, \begin{array}{c} 4 \\ 4 \\ 7 \end{array} \right\}$ Contributed by Robert Beezer Solution [349]



C30 For the matrix B below, find a set S that is linearly independent and spans the null

space of B, that is, $\mathcal{N}(B) = \langle S \rangle$.

$$B = \begin{bmatrix} -3 & 1 & -2 & 7\\ -1 & 2 & 1 & 4\\ 1 & 1 & 2 & -1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [351]

C31 For the matrix A below, find a linearly independent set S so that the null space of A is spanned by S, that is, $\mathcal{N}(A) = \langle S \rangle$.

$$A = \begin{bmatrix} -1 & -2 & 2 & 1 & 5\\ 1 & 2 & 1 & 1 & 5\\ 3 & 6 & 1 & 2 & 7\\ 2 & 4 & 0 & 1 & 2 \end{bmatrix}$$

Contributed by Robert Beezer Solution [352]

C32 Find a set of column vectors, T, such that (1) the span of T is the null space of B, $\langle T \rangle = \mathcal{N}(B)$ and (2) T is a linearly independent set.

$$B = \begin{bmatrix} 2 & 1 & 1 & 1 \\ -4 & -3 & 1 & -7 \\ 1 & 1 & -1 & 3 \end{bmatrix}$$

Contributed by Robert Beezer Solution [353]

C33 Find a set S so that S is linearly independent and $\mathcal{N}(A) = \langle S \rangle$, where $\mathcal{N}(A)$ is the null space of the matrix A below.

	2	3	3	1	4]
A =	1	1	-1	-1	-3
A =	3	2	-8	-1	1

Contributed by Robert Beezer Solution [354]

C50 Consider each archetype that is a system of equations and consider the solutions listed for the homogeneous version of the archetype. (If only the trivial solution is listed, then assume this is the only solution to the system.) From the solution set, determine if the columns of the coefficient matrix form a linearly independent or linearly dependent set. In the case of a linearly dependent set, use one of the sample solutions to provide a nontrivial relation of linear dependence on the set of columns of the coefficient matrix (Definition RLD [735]). Indicate when Theorem MVSLD [328] applies and connect this with the number of variables and equations in the system of equations.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678]

```
Archetype F [1686]
Archetype G [1697]/Archetype H [1705]
Archetype I [1714]
Archetype J [1723]
```

Contributed by Robert Beezer

C51 For each archetype that is a system of equations consider the homogeneous version. Write elements of the solution set in vector form (Theorem VFSLS [244]) and from this extract the vectors \mathbf{z}_j described in Theorem BNS [334]. These vectors are used in a span construction to describe the null space of the coefficient matrix for each archetype. What does it mean when we write a null space as $\langle \{ \} \rangle$?

```
Archetype A [1642]
Archetype B [1652]
Archetype C [1662]
Archetype D [1670]/Archetype E [1678]
Archetype F [1686]
Archetype G [1697]/Archetype H [1705]
Archetype I [1714]
Archetype J [1723]
```

Contributed by Robert Beezer

C52 For each archetype that is a system of equations consider the homogeneous version.

Sample solutions are given and a linearly independent spanning set is given for the null space of the coefficient matrix. Write each of the sample solutions individually as a linear combination of the vectors in the spanning set for the null space of the coefficient matrix.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678] Archetype F [1686] Archetype G [1697]/Archetype H [1705] Archetype I [1714] Archetype J [1723]

Contributed by Robert Beezer

C60 For the matrix A below, find a set of vectors S so that (1) S is linearly independent, and (2) the span of S equals the null space of A, $\langle S \rangle = \mathcal{N}(A)$. (See Exercise SS.C60 [298].)

$$A = \begin{bmatrix} 1 & 1 & 6 & -8\\ 1 & -2 & 0 & 1\\ -2 & 1 & -6 & 7 \end{bmatrix}$$

Contributed by Robert Beezer Solution [355]

M20 Suppose that $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3}$ is a set of three vectors from \mathbb{C}^{873} . Prove that the set

$$T = \{2\mathbf{v}_1 + 3\mathbf{v}_2 + \mathbf{v}_3, \, \mathbf{v}_1 - \mathbf{v}_2 - 2\mathbf{v}_3, \, 2\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3\}$$

is linearly dependent.

Contributed by Robert Beezer Solution [356]

M21 Suppose that $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3}$ is a linearly independent set of three vectors from \mathbb{C}^{873} . Prove that the set

$$T = \{2\mathbf{v}_1 + 3\mathbf{v}_2 + \mathbf{v}_3, \, \mathbf{v}_1 - \mathbf{v}_2 + 2\mathbf{v}_3, \, 2\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3\}$$

is linearly independent.

Contributed by Robert Beezer Solution [357]

M50 Consider the set of vectors from \mathbb{C}^3 , W, given below. Find a set T that contains three vectors from W and such that $W = \langle T \rangle$.

$$W = \langle \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4, \, \mathbf{v}_5\} \rangle = \left\langle \left\{ \begin{bmatrix} 2\\1\\1\\1 \end{bmatrix}, \, \begin{bmatrix} -1\\-1\\1\\1 \end{bmatrix}, \, \begin{bmatrix} 1\\2\\3\\3 \end{bmatrix}, \, \begin{bmatrix} 3\\1\\3\\3 \end{bmatrix}, \, \begin{bmatrix} 0\\1\\-3\\3 \end{bmatrix} \right\} \right\rangle$$

Contributed by Robert Beezer Solution [358]

M51 Consider the subspace $W = \langle \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \rangle$. Find a set S so that (1) S is a subset of W, (2) S is linearly independent, and (3) $W = \langle S \rangle$. Write each vector not included in S

as a linear combination of the vectors that are in S.

$$\mathbf{v}_1 = \begin{bmatrix} 1\\ -1\\ 2 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} 4\\ -4\\ 8 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} -3\\ 2\\ -7 \end{bmatrix} \qquad \mathbf{v}_4 = \begin{bmatrix} 2\\ 1\\ 7 \end{bmatrix}$$

Contributed by Manley Perkel Solution [359]

T10 Prove that if a set of vectors contains the zero vector, then the set is linearly dependent. (Ed. "The zero vector is death to linearly independent sets.") Contributed by Martin Jackson

T12 Suppose that S is a linearly independent set of vectors, and T is a subset of $S, T \subseteq S$ (Definition SSET [1603]). Prove that T is linearly independent. Contributed by Robert Beezer

T13 Suppose that T is a linearly dependent set of vectors, and T is a subset of $S, T \subseteq S$ (Definition SSET [1603]). Prove that S is linearly dependent. Contributed by Robert Beezer

T15 Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n\}$ is a set of vectors. Prove that

$$\{\mathbf{v}_1 - \mathbf{v}_2, \, \mathbf{v}_2 - \mathbf{v}_3, \, \mathbf{v}_3 - \mathbf{v}_4, \, \dots, \, \mathbf{v}_n - \mathbf{v}_1\}$$

is a linearly dependent set.

Contributed by Robert Beezer Solution [360]

T20 Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is a linearly independent set in \mathbb{C}^{35} . Prove that

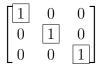
$$\{\mathbf{v}_1,\,\mathbf{v}_1+\mathbf{v}_2,\,\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3,\,\mathbf{v}_1+\mathbf{v}_2+\mathbf{v}_3+\mathbf{v}_4\}$$

is a linearly independent set. Contributed by Robert Beezer Solution [361]

T50 Suppose that A is an $m \times n$ matrix with linearly independent columns and the linear system $\mathcal{LS}(A, \mathbf{b})$ is consistent. Show that this system has a unique solution. (Notice that we are not requiring A to be square.) Contributed by Robert Beezer Solution [362]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [339] With three vectors from \mathbb{C}^3 , we can form a square matrix by making these three vectors the columns of a matrix. We do so, and row-reduce to obtain,



the 3×3 identity matrix. So by Theorem NME2 [330] the original matrix is nonsingular and its columns are therefore a linearly independent set.

C21 Contributed by Robert Beezer Statement [339] Theorem LIVRN [325] says we can answer this question by putting theses vectors into a matrix as columns and row-reducing. Doing this we obtain,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

With n = 3 (3 vectors, 3 columns) and r = 3 (3 leading 1's) we have n = r and the theorem says the vectors are linearly independent.

C22 Contributed by Robert Beezer Statement [339] Five vectors from \mathbb{C}^3 . Theorem MVSLD [328] says the set is linearly dependent. Boom.

C23 Contributed by Robert Beezer Statement [340] Theorem LIVRN [325] suggests we analyze a matrix whose columns are the vectors of S,

$$A = \begin{bmatrix} 1 & 3 & 2 & 1 \\ -2 & 3 & 1 & 0 \\ 2 & 1 & 2 & 1 \\ 5 & 2 & -1 & 2 \\ 3 & -4 & 1 & 2 \end{bmatrix}$$

Row-reducing the matrix A yields,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We see that r = 4 = n, where r is the number of nonzero rows and n is the number of columns. By Theorem LIVRN [325], the set S is linearly independent.

C24 Contributed by Robert Beezer Statement [340]

Theorem LIVRN [325] suggests we analyze a matrix whose columns are the vectors from the set,

$$A = \begin{bmatrix} 1 & 3 & 4 & -1 \\ 2 & 2 & 4 & 2 \\ -1 & -1 & -2 & -1 \\ 0 & 2 & 2 & -2 \\ 1 & 2 & 3 & 0 \end{bmatrix}$$

Row-reducing the matrix A yields,

We see that $r = 2 \neq 4 = n$, where r is the number of nonzero rows and n is the number of columns. By Theorem LIVRN [325], the set S is linearly dependent.

C25 Contributed by Robert Beezer Statement [340] Theorem LIVRN [325] suggests we analyze a matrix whose columns are the vectors from the set,

$$A = \begin{bmatrix} 2 & 4 & 10\\ 1 & -2 & -7\\ 3 & 1 & 0\\ -1 & 3 & 10\\ 2 & 2 & 4 \end{bmatrix}$$

Row-reducing the matrix A yields,

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

We see that $r = 2 \neq 3 = n$, where r is the number of nonzero rows and n is the number of columns. By Theorem LIVRN [325], the set S is linearly dependent.

C30 Contributed by Robert Beezer Statement [340]

The requested set is described by Theorem BNS [334]. It is easiest to find by using the procedure of Example VFSAL [252]. Begin by row-reducing the matrix, viewing it as the coefficient matrix of a homogeneous system of equations. We obtain,

$$\begin{bmatrix} 1 & 0 & 1 & -2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Now build the vector form of the solutions to this homogeneous system (Theorem VFSLS [244]). The free variables are x_3 and x_4 , corresponding to the columns without leading 1's,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

The desired set S is simply the constant vectors in this expression, and these are the vectors \mathbf{z}_1 and \mathbf{z}_2 described by Theorem BNS [334].

$$S = \left\{ \begin{bmatrix} -1\\ -1\\ 1\\ 0 \end{bmatrix}, \begin{bmatrix} 2\\ -1\\ 0\\ 1 \end{bmatrix} \right\}$$

C31 Contributed by Robert Beezer Statement [341]

Theorem BNS [334] provides formulas for n - r vectors that will meet the requirements of this question. These vectors are the same ones listed in Theorem VFSLS [244] when we solve the homogeneous system $\mathcal{LS}(A, \mathbf{0})$, whose solution set is the null space (Definition NSM [152]).

To apply Theorem BNS [334] or Theorem VFSLS [244] we first row-reduce the matrix,

resulting in

$$B = \begin{bmatrix} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 6 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So we see that n - r = 5 - 3 = 2 and $F = \{2, 5\}$, so the vector form of a generic solution vector is

$\begin{bmatrix} x_1 \end{bmatrix}$		$\begin{bmatrix} -2 \end{bmatrix}$		[-3]
x_2		1		0
x_3	$=x_2$	0	$+x_{5}$	-6
x_4		0		4
x_5		0		1

So we have

$$\mathcal{N}(A) = \left\langle \left\{ \begin{bmatrix} -2\\1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} -3\\0\\-6\\4\\1 \end{bmatrix} \right\} \right\rangle$$

C32 Contributed by Robert Beezer Statement [341]

The conclusion of Theorem BNS [334] gives us everything this question asks for. We need the reduced row-echelon form of the matrix so we can determine the number of vectors in

T, and their entries.

$$\begin{bmatrix} 2 & 1 & 1 & 1 \\ -4 & -3 & 1 & -7 \\ 1 & 1 & -1 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & -3 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We can build the set T in immediately via Theorem BNS [334], but we will illustrate its construction in two steps. Since $F = \{3, 4\}$, we will have two vectors and can distribute strategically placed ones, and many zeros. Then we distribute the negatives of the appropriate entries of the non-pivot columns of the reduced row-echelon matrix.

$$T = \left\{ \begin{bmatrix} 1\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\1 \end{bmatrix} \right\} \qquad \qquad T = \left\{ \begin{bmatrix} -2\\3\\1\\0 \end{bmatrix}, \begin{bmatrix} 2\\-5\\0\\1 \end{bmatrix} \right\}$$

C33 Contributed by Robert Beezer Statement [342] A direct application of Theorem BNS [334] will provide the desired set. We require the reduced row-echelon form of A.

$$\begin{bmatrix} 2 & 3 & 3 & 1 & 4 \\ 1 & 1 & -1 & -1 & -3 \\ 3 & 2 & -8 & -1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -6 & 0 & 3 \\ 0 & 1 & 5 & 0 & -2 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

The non-pivot columns have indices $F = \{3, 5\}$. We build the desired set in two steps, first placing the requisite zeros and ones in locations based on F, then placing the negatives of the entries of columns 3 and 5 in the proper locations. This is all specified in Theorem BNS [334].

$$S = \left\{ \begin{bmatrix} 1\\ 1\\ 0 \end{bmatrix}, \begin{bmatrix} 0\\ 1 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} 6\\ -5\\ 1\\ 0\\ 0 \end{bmatrix}, \begin{bmatrix} -3\\ 2\\ 0\\ -4\\ 1 \end{bmatrix} \right\}$$

C60 Contributed by Robert Beezer Statement [344]

Theorem BNS [334] says that if we find the vector form of the solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$, then the fixed vectors (one per free variable) will have the desired properties. Row-reduce A, viewing it as the augmented matrix of a homogeneous system with an invisible columns of zeros as the last column,

$$\begin{bmatrix} 1 & 0 & 4 & -5 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Moving to the vector form of the solutions (Theorem VFSLS [244]), with free variables x_3 and x_4 , solutions to the consistent system (it is homogeneous, Theorem HSC [147]) can be

expressed as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} -4 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 5 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Then with S given by

$$S = \left\{ \begin{bmatrix} -4\\-2\\1\\0 \end{bmatrix}, \begin{bmatrix} 5\\3\\0\\1 \end{bmatrix} \right\}$$

Theorem BNS [334] guarantees the set has the desired properties.

M20 Contributed by Robert Beezer Statement [345] By Definition LICV [317], we can complete this problem by finding scalars, α_1 , α_2 , α_3 , not all zero, such that

$$\alpha_1 (2\mathbf{v}_1 + 3\mathbf{v}_2 + \mathbf{v}_3) + \alpha_2 (\mathbf{v}_1 - \mathbf{v}_2 - 2\mathbf{v}_3) + \alpha_3 (2\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3) = \mathbf{0}$$

Using various properties in Theorem VSPCV [205], we can rearrange this vector equation to

$$(2\alpha_1 + \alpha_2 + 2\alpha_3) \mathbf{v}_1 + (3\alpha_1 - \alpha_2 + \alpha_3) \mathbf{v}_2 + (\alpha_1 - 2\alpha_2 - \alpha_3) \mathbf{v}_3 = \mathbf{0}$$

We can certainly make this vector equation true if we can determine values for the α 's such that

$$2\alpha_1 + \alpha_2 + 2\alpha_3 = 0$$

Version 2.99

$$3\alpha_1 - \alpha_2 + \alpha_3 = 0$$
$$\alpha_1 - 2\alpha_2 - \alpha_3 = 0$$

Aah, a homogeneous system of equations. And it has infinitely many non-zero solutions. By the now familiar techniques, one such solution is $\alpha_1 = 3$, $\alpha_2 = 4$, $\alpha_3 = -5$, which you can check in the original relation of linear dependence on T above.

Note that simply writing down the three scalars, and demonstrating that they provide a nontrivial relation of linear dependence on T, could be considered an ironclad solution. But it wouldn't have been very informative for you if we had only done just that here. Compare this solution very carefully with Solution LI.M21 [357].

M21 Contributed by Robert Beezer Statement [345] By Definition LICV [317] we can complete this problem by proving that if we assume that

$$\alpha_1 (2\mathbf{v}_1 + 3\mathbf{v}_2 + \mathbf{v}_3) + \alpha_2 (\mathbf{v}_1 - \mathbf{v}_2 + 2\mathbf{v}_3) + \alpha_3 (2\mathbf{v}_1 + \mathbf{v}_2 - \mathbf{v}_3) = \mathbf{0}$$

then we must conclude that $\alpha_1 = \alpha_2 = \alpha_3 = 0$. Using various properties in Theorem VSPCV [205], we can rearrange this vector equation to

$$(2\alpha_1 + \alpha_2 + 2\alpha_3)\mathbf{v}_1 + (3\alpha_1 - \alpha_2 + \alpha_3)\mathbf{v}_2 + (\alpha_1 + 2\alpha_2 - \alpha_3)\mathbf{v}_3 = \mathbf{0}$$

Because the set $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3}$ was assumed to be linearly independent, by Definition LICV [317] we must conclude that

$$2\alpha_1 + \alpha_2 + 2\alpha_3 = 0$$

 $3\alpha_1 - \alpha_2 + \alpha_3 = 0$ $\alpha_1 + 2\alpha_2 - \alpha_3 = 0$

Aah, a homogeneous system of equations. And it has a unique solution, the trivial solution. So, $\alpha_1 = \alpha_2 = \alpha_3 = 0$, as desired. It is an inescapable conclusion from our assumption of a relation of linear dependence above. Done.

Compare this solution very carefully with Solution LI.M20 [356], noting especially how this problem required (and used) the hypothesis that the original set be linearly independent, and how this solution feels more like a proof, while the previous problem could be solved with a fairly simple demonstration of any nontrivial relation of linear dependence.

M50 Contributed by Robert Beezer Statement [345]

We want to first find some relations of linear dependence on { \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 , \mathbf{v}_4 , \mathbf{v}_5 } that will allow us to "kick out" some vectors, in the spirit of Example SCAD [290]. To find relations of linear dependence, we formulate a matrix A whose columns are \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 , \mathbf{v}_4 , \mathbf{v}_5 . Then we consider the homogeneous system of equations $\mathcal{LS}(A, \mathbf{0})$ by row-reducing its coefficient matrix (remember that if we formulated the augmented matrix we would just add a column of zeros). After row-reducing, we obtain

$$\begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

From this we find that solutions can be obtained employing the free variables x_4 and x_5 . With appropriate choices we will be able to conclude that vectors \mathbf{v}_4 and \mathbf{v}_5 are unnecessary for creating W via a span. By Theorem SLSLC [229] the choice of free variables below lead to solutions and linear combinations, which are then rearranged.

$$\begin{aligned} x_4 &= 1, x_5 = 0 \Rightarrow (-2)\mathbf{v}_1 + (-1)\mathbf{v}_2 + (0)\mathbf{v}_3 + (1)\mathbf{v}_4 + (0)\mathbf{v}_5 = \mathbf{0} \Rightarrow \mathbf{v}_4 = 2\mathbf{v}_1 + \mathbf{v}_2 \\ x_4 &= 0, x_5 = 1 \Rightarrow (1)\mathbf{v}_1 + (2)\mathbf{v}_2 + (0)\mathbf{v}_3 + (0)\mathbf{v}_4 + (1)\mathbf{v}_5 = \mathbf{0} \Rightarrow \mathbf{v}_5 = -\mathbf{v}_1 - 2\mathbf{v}_2 \end{aligned}$$

Since \mathbf{v}_4 and \mathbf{v}_5 can be expressed as linear combinations of \mathbf{v}_1 and \mathbf{v}_2 we can say that \mathbf{v}_4 and \mathbf{v}_5 are not needed for the linear combinations used to build W (a claim that we could establish carefully with a pair of set equality arguments). Thus

$$W = \left\langle \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 2\\1\\1\\1 \end{bmatrix}, \, \begin{bmatrix} -1\\-1\\1\\1 \end{bmatrix}, \, \begin{bmatrix} 1\\2\\3\\3 \end{bmatrix} \right\} \right\rangle$$

That the $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent set can be established quickly with Theorem LIVRN [325].

There are other answers to this question, but notice that any nontrivial linear combination of \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 , \mathbf{v}_4 , \mathbf{v}_5 will have a zero coefficient on \mathbf{v}_3 , so this vector can never be eliminated from the set used to build the span.

M51 Contributed by Robert Beezer Statement [345]

This problem can be solved using the approach in Solution LI.M50 [358]. We will provide a solution here that is more ad-hoc, but note that we will have a more straight-forward procedure given by the upcoming Theorem BS [376].

 \mathbf{v}_1 is a non-zero vector, so in a set all by itself we have a linearly independent set. As \mathbf{v}_2 is a scalar multiple of \mathbf{v}_1 , the equation $-4\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$ is a relation of linear dependence on $\{\mathbf{v}_1, \mathbf{v}_2\}$, so we will pass on \mathbf{v}_2 . No such relation of linear dependence exists on $\{\mathbf{v}_1, \mathbf{v}_3\}$, though on $\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4\}$ we have the relation of linear dependence $7\mathbf{v}_1 + 3\mathbf{v}_3 + \mathbf{v}_4 = \mathbf{0}$. So take $S = \{\mathbf{v}_1, \mathbf{v}_3\}$, which is linearly independent.

Then

$$v_2 = 4v_1 + 0v_3$$
 $v_4 = -7v_1 - 3v_3$

The two equations above are enough to justify the set equality

$$W = \langle \{ \mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4 \} \rangle = \langle \{ \mathbf{v}_1, \, \mathbf{v}_3 \} \rangle = \langle S \rangle$$

There are other solutions (for example, swap the roles of \mathbf{v}_1 and \mathbf{v}_2 , but by upcoming theorems we can confidently claim that any solution will be a set S with exactly two vectors.

T15 Contributed by Robert Beezer Statement [346] Consider the following linear combination

$$1 (\mathbf{v}_{1} - \mathbf{v}_{2}) + 1 (\mathbf{v}_{2} - \mathbf{v}_{3}) + 1 (\mathbf{v}_{3} - \mathbf{v}_{4}) + \dots + 1 (\mathbf{v}_{n} - \mathbf{v}_{1})$$

= $\mathbf{v}_{1} - \mathbf{v}_{2} + \mathbf{v}_{2} - \mathbf{v}_{3} + \mathbf{v}_{3} - \mathbf{v}_{4} + \dots + \mathbf{v}_{n} - \mathbf{v}_{1}$
= $\mathbf{v}_{1} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0} - \mathbf{v}_{1}$
= $\mathbf{0}$

This is a nontrivial relation of linear dependence (Definition RLDCV [316]), so by Definition LICV [317] the set is linearly dependent.

T20 Contributed by Robert Beezer Statement [347]

Our hypothesis and our conclusion use the term linear independence, so it will get a workout. To establish linear independence, we begin with the definition (Definition LICV [317]) and write a relation of linear dependence (Definition RLDCV [316]),

$$\alpha_{1}(\mathbf{v}_{1}) + \alpha_{2}(\mathbf{v}_{1} + \mathbf{v}_{2}) + \alpha_{3}(\mathbf{v}_{1} + \mathbf{v}_{2} + \mathbf{v}_{3}) + \alpha_{4}(\mathbf{v}_{1} + \mathbf{v}_{2} + \mathbf{v}_{3} + \mathbf{v}_{4}) = \mathbf{0}$$

Using the distributive and commutative properties of vector addition and scalar multiplication (Theorem VSPCV [205]) this equation can be rearranged as

 $(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) \mathbf{v}_1 + (\alpha_2 + \alpha_3 + \alpha_4) \mathbf{v}_2 + (\alpha_3 + \alpha_4) \mathbf{v}_3 + (\alpha_4) \mathbf{v}_4 = \mathbf{0}$

However, this is a relation of linear dependence (Definition RLDCV [316]) on a linearly independent set, $\{v_1, v_2, v_3, v_4\}$ (this was our lone hypothesis). By the definition of linear independence (Definition LICV [317]) the scalars must all be zero. This is the homogeneous system of equations,

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0$$
$$\alpha_2 + \alpha_3 + \alpha_4 = 0$$
$$\alpha_3 + \alpha_4 = 0$$
$$\alpha_4 = 0$$

Row-reducing the coefficient matrix of this system (or backsolving) gives the conclusion

$$\alpha_1 = 0 \qquad \qquad \alpha_2 = 0 \qquad \qquad \alpha_3 = 0 \qquad \qquad \alpha_4 = 0$$

This means, by Definition LICV [317], that the original set

$$\{\mathbf{v}_1, \, \mathbf{v}_1 + \mathbf{v}_2, \, \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3, \, \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4\}$$

is linearly independent.

T50 Contributed by Robert Beezer Statement [347]

Let $A = [\mathbf{A}_1 | \mathbf{A}_2 | \mathbf{A}_3 | \dots | \mathbf{A}_n]$. $\mathcal{LS}(A, \mathbf{b})$ is consistent, so we know the system has at least one solution (Definition CS [113]). We would like to show that there are no more than one solution to the system. Employing Technique U [1625], suppose that \mathbf{x} and \mathbf{y} are two solution vectors for $\mathcal{LS}(A, \mathbf{b})$. By Theorem SLSLC [229] we know we can write,

$$\mathbf{b} = [\mathbf{x}]_1 A_1 + [\mathbf{x}]_2 A_2 + [\mathbf{x}]_3 A_3 + \dots + [\mathbf{x}]_n A_n$$

$$\mathbf{b} = [\mathbf{y}]_1 A_1 + [\mathbf{y}]_2 A_2 + [\mathbf{y}]_3 A_3 + \dots + [\mathbf{y}]_n A_n$$

Then

$$0 = \mathbf{b} - \mathbf{b}$$

= $([\mathbf{x}]_1 A_1 + [\mathbf{x}]_2 A_2 + \dots + [\mathbf{x}]_n A_n) - ([\mathbf{y}]_1 A_1 + [\mathbf{y}]_2 A_2 + \dots + [\mathbf{y}]_n A_n)$
= $([\mathbf{x}]_1 - [\mathbf{y}]_1) A_1 + ([\mathbf{x}]_2 - [\mathbf{y}]_2) A_2 + \dots + ([\mathbf{x}]_n - [\mathbf{y}]_n) A_n$

This is a relation of linear dependence (Definition RLDCV [316]) on a linearly independent set (the columns of A). So the scalars *must* all be zero,

$$[\mathbf{x}]_1 - [\mathbf{y}]_1 = 0$$
 $[\mathbf{x}]_2 - [\mathbf{y}]_2 = 0$... $[\mathbf{x}]_n - [\mathbf{y}]_n = 0$

Version 2.99

Rearranging these equations yields the statement that $[\mathbf{x}]_i = [\mathbf{y}]_i$, for $1 \le i \le n$. However, this is exactly how we define vector equality (Definition CVE [199]), so $\mathbf{x} = \mathbf{y}$ and the system has only one solution.

Section LDS Linear Dependence and Spans

In any linearly dependent set there is always one vector that can be written as a linear combination of the others. This is the substance of the upcoming Theorem DLDS [365]. Perhaps this will explain the use of the word "dependent." In a linearly dependent set, at least one vector "depends" on the others (via a linear combination).

Indeed, because Theorem DLDS [365] is an equivalence (Technique E [1619]) some authors use this condition as a definition (Technique D [1611]) of linear dependence. Then linear independence is defined as the logical opposite of linear dependence. Of course, we have *chosen* to take Definition LICV [317] as our definition, and then follow with Theorem DLDS [365] as a theorem.

Subsection LDSS Linearly Dependent Sets and Spans

If we use a linearly dependent set to construct a span, then we can *always* create the same infinite set with a starting set that is one vector smaller in size. We will illustrate this

behavior in Example RSC5 [366]. However, this will not be possible if we build a span from a linearly independent set. So in a certain sense, using a linearly independent set to formulate a span is the best possible way — there aren't any extra vectors being used to build up all the necessary linear combinations. OK, here's the theorem, and then the example.

Theorem DLDS

Dependency in Linearly Dependent Sets

Suppose that $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is a set of vectors. Then S is a linearly dependent set if and only if there is an index $t, 1 \le t \le n$ such that \mathbf{u}_t is a linear combination of the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_{t-1}, \mathbf{u}_{t+1}, \dots, \mathbf{u}_n$.

Proof (\Rightarrow) Suppose that S is linearly dependent, so there exists a nontrivial relation of linear dependence by Definition LICV [317]. That is, there are scalars, α_i , $1 \le i \le n$, which are not all zero, such that

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \dots + \alpha_n\mathbf{u}_n = \mathbf{0}.$$

Since the α_i cannot all be zero, choose one, say α_t , that is nonzero. Then,

$$\mathbf{u}_{t} = \frac{-1}{\alpha_{t}} (-\alpha_{t} \mathbf{u}_{t})$$
Property MICN [1598]
$$= \frac{-1}{\alpha_{t}} (\alpha_{1} \mathbf{u}_{1} + \dots + \alpha_{t-1} \mathbf{u}_{t-1} + \alpha_{t+1} \mathbf{u}_{t+1} + \dots + \alpha_{n} \mathbf{u}_{n})$$
Theorem VSPCV [205]
$$= \frac{-\alpha_{1}}{\alpha_{t}} \mathbf{u}_{1} + \dots + \frac{-\alpha_{t-1}}{\alpha_{t}} \mathbf{u}_{t-1} + \frac{-\alpha_{t+1}}{\alpha_{t}} \mathbf{u}_{t+1} + \dots + \frac{-\alpha_{n}}{\alpha_{t}} \mathbf{u}_{n}$$
Theorem VSPCV [205]

Since the values of $\frac{\alpha_i}{\alpha_t}$ are again scalars, we have expressed \mathbf{u}_t as a linear combination of the other elements of S.

(\Leftarrow) Assume that the vector \mathbf{u}_t is a linear combination of the other vectors in S. Write this linear combination, denoting the relevant scalars as $\beta_1, \beta_2, \ldots, \beta_{t-1}, \beta_{t+1}, \ldots, \beta_n$, as

$$\mathbf{u}_{\mathbf{t}} = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \dots + \beta_{t-1} \mathbf{u}_{t-1} + \beta_{t+1} \mathbf{u}_{t+1} + \dots + \beta_n \mathbf{u}_n$$

Then we have

$$\beta_{1}\mathbf{u}_{1} + \dots + \beta_{t-1}\mathbf{u}_{t-1} + (-1)\mathbf{u}_{t} + \beta_{t+1}\mathbf{u}_{t+1} + \dots + \beta_{n}\mathbf{u}_{n}$$

$$= \mathbf{u}_{t} + (-1)\mathbf{u}_{t}$$
Theorem VSPCV [205]
$$= (1 + (-1))\mathbf{u}_{t}$$
Property DSAC [206]
$$= 0\mathbf{u}_{t}$$
Property AICN [1597]
$$= \mathbf{0}$$
Definition CVSM [203]

So the scalars $\beta_1, \beta_2, \beta_3, \ldots, \beta_{t-1}, \beta_t = -1, \beta_{t+1}, \ldots, \beta_n$ provide a *nontrivial* linear combination of the vectors in S, thus establishing that S is a linearly dependent set (Definition LICV [317]).

This theorem can be used, sometimes repeatedly, to whittle down the size of a set of vectors used in a span construction. We have seen some of this already in Example SCAD [290], but in the next example we will detail some of the subtleties.

Example RSC5 Reducing a span in \mathbb{C}^5

Consider the set of n = 4 vectors from \mathbb{C}^5 ,

$$R = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4\} = \left\{ \begin{bmatrix} 1\\2\\-1\\3\\2 \end{bmatrix}, \begin{bmatrix} 2\\1\\3\\1\\2 \end{bmatrix}, \begin{bmatrix} 0\\-7\\6\\-11\\-2 \end{bmatrix}, \begin{bmatrix} 4\\1\\2\\1\\6 \end{bmatrix} \right\}$$

and define $V = \langle R \rangle$.

To employ Theorem LIVHS [321], we form a 5×4 coefficient matrix, D,

$$D = \begin{bmatrix} 1 & 2 & 0 & 4 \\ 2 & 1 & -7 & 1 \\ -1 & 3 & 6 & 2 \\ 3 & 1 & -11 & 1 \\ 2 & 2 & -2 & 6 \end{bmatrix}$$

and row-reduce to understand solutions to the homogeneous system $\mathcal{LS}(D, \mathbf{0})$,

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We can find infinitely many solutions to this system, most of them nontrivial, and we choose any one we like to build a relation of linear dependence on R. Let's begin with $x_4 = 1$, to find the solution

 $\begin{vmatrix} \mathbf{u} \\ \mathbf{0} \\ -\mathbf{1} \end{vmatrix}$

So we can write the relation of linear dependence,

$$(-4)\mathbf{v}_1 + 0\mathbf{v}_2 + (-1)\mathbf{v}_3 + 1\mathbf{v}_4 = \mathbf{0}$$

Theorem DLDS [365] guarantees that we can solve this relation of linear dependence for *some* vector in R, but the choice of which one is up to us. Notice however that \mathbf{v}_2 has a zero coefficient. In this case, we cannot choose to solve for \mathbf{v}_2 . Maybe some other relation of linear dependence would produce a nonzero coefficient for \mathbf{v}_2 if we just had to solve for this vector. Unfortunately, this example has been engineered to *always* produce a zero coefficient here, as you can see from solving the homogeneous system. Every solution has $x_2 = 0$!

OK, if we are convinced that we cannot solve for \mathbf{v}_2 , let's instead solve for \mathbf{v}_3 ,

$$\mathbf{v}_3 = (-4)\mathbf{v}_1 + 0\mathbf{v}_2 + 1\mathbf{v}_4 = (-4)\mathbf{v}_1 + 1\mathbf{v}_4$$

We now claim that this particular equation will allow us to write

$$V = \langle R \rangle = \langle \{ \mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4 \} \rangle = \langle \{ \mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_4 \} \rangle$$

in essence declaring \mathbf{v}_3 as surplus for the task of building V as a span. This claim is an equality of two sets, so we will use Definition SE [1604] to establish it carefully. Let $R' = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4}$ and $V' = \langle R' \rangle$. We want to show that V = V'.

First show that $V' \subseteq V$. Since every vector of R' is in R, any vector we can construct in V' as a linear combination of vectors from R' can also be constructed as a vector in V by the same linear combination of the same vectors in R. That was easy, now turn it around.

Next show that $V \subseteq V'$. Choose any **v** from V. Then there are scalars $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ so that

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 \mathbf{v}_3 + \alpha_4 \mathbf{v}_4$$

= $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \alpha_3 ((-4)\mathbf{v}_1 + 1\mathbf{v}_4) + \alpha_4 \mathbf{v}_4$
= $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + ((-4\alpha_3)\mathbf{v}_1 + \alpha_3 \mathbf{v}_4) + \alpha_4 \mathbf{v}_4$
= $(\alpha_1 - 4\alpha_3)\mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + (\alpha_3 + \alpha_4)\mathbf{v}_4.$

This equation says that \mathbf{v} can then be written as a linear combination of the vectors in R'and hence qualifies for membership in V'. So $V \subseteq V'$ and we have established that V = V'.

If R' was also linearly dependent (it is not), we could reduce the set even further. Notice that we could have chosen to eliminate any one of \mathbf{v}_1 , \mathbf{v}_3 or \mathbf{v}_4 , but somehow \mathbf{v}_2 is essential to the creation of V since it cannot be replaced by any linear combination of \mathbf{v}_1 , \mathbf{v}_3 or \mathbf{v}_4 .

Subsection COV Casting Out Vectors

In Example RSC5 [366] we used four vectors to create a span. With a relation of linear dependence in hand, we were able to "toss out" one of these four vectors and create the same span from a subset of just three vectors from the original set of four. We did have to take some care as to just which vector we tossed out. In the next example, we will be more methodical about just how we choose to eliminate vectors from a linearly dependent set while preserving a span.

Example COV Casting out vectors

We begin with a set S containing seven vectors from \mathbb{C}^4 ,

$$S = \left\{ \begin{bmatrix} 1\\2\\0\\-1 \end{bmatrix}, \begin{bmatrix} 4\\8\\0\\-4 \end{bmatrix}, \begin{bmatrix} 0\\-1\\2\\2 \end{bmatrix}, \begin{bmatrix} -1\\3\\-3\\4 \end{bmatrix}, \begin{bmatrix} 0\\9\\-4\\8 \end{bmatrix}, \begin{bmatrix} 7\\-13\\12\\-31 \end{bmatrix}, \begin{bmatrix} -9\\7\\-8\\37 \end{bmatrix} \right\}$$

and define $W = \langle S \rangle$. The set S is obviously linearly dependent by Theorem MVSLD [328], since we have n = 7 vectors from \mathbb{C}^4 . So we can slim down S some, and still create W as the span of a smaller set of vectors. As a device for identifying relations of linear dependence among the vectors of S, we place the seven column vectors of S into a matrix as columns,

$$A = [\mathbf{A}_1 | \mathbf{A}_2 | \mathbf{A}_3 | \dots | \mathbf{A}_7] = \begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 \\ 0 & 0 & 2 & -3 & -4 & 12 & -8 \\ -1 & -4 & 2 & 4 & 8 & -31 & 37 \end{bmatrix}$$

By Theorem SLSLC [229] a nontrivial solution to $\mathcal{LS}(A, \mathbf{0})$ will give us a nontrivial relation of linear dependence (Definition RLDCV [316]) on the columns of A (which are the elements of the set S). The row-reduced form for A is the matrix

$$B = \begin{bmatrix} 1 & 4 & 0 & 0 & 2 & 1 & -3 \\ 0 & 0 & 1 & 0 & 1 & -3 & 5 \\ 0 & 0 & 0 & 1 & 2 & -6 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

so we can easily create solutions to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ using the free variables x_2, x_5, x_6, x_7 . Any such solution will correspond to a relation of linear dependence on the columns of B. These solutions will allow us to solve for one column vector as a linear combination of some others, in the spirit of Theorem DLDS [365], and remove that vector from the set. We'll set about forming these linear combinations methodically. Set the free

variable x_2 to one, and set the other free variables to zero. Then a solution to $\mathcal{LS}(A, \mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} -4\\1\\0\\0\\0\\0\\0\end{bmatrix}$$

which can be used to create the linear combination

$$(-4)\mathbf{A}_1 + 1\mathbf{A}_2 + 0\mathbf{A}_3 + 0\mathbf{A}_4 + 0\mathbf{A}_5 + 0\mathbf{A}_6 + 0\mathbf{A}_7 = \mathbf{0}$$

This can then be arranged and solved for A_2 , resulting in A_2 expressed as a linear combination of $\{A_1, A_3, A_4\}$,

$$\mathbf{A}_2 = 4\mathbf{A}_1 + 0\mathbf{A}_3 + 0\mathbf{A}_4$$

This means that \mathbf{A}_2 is surplus, and we can create W just as well with a smaller set with this vector removed,

$$W = \langle \{\mathbf{A}_1, \, \mathbf{A}_3, \, \mathbf{A}_4, \, \mathbf{A}_5, \, \mathbf{A}_6, \, \mathbf{A}_7 \} \rangle$$

Technically, this set equality for W requires a proof, in the spirit of Example RSC5 [366], but we will bypass this requirement here, and in the next few paragraphs.

Now, set the free variable x_5 to one, and set the other free variables to zero. Then a solution to $\mathcal{LS}(B, \mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} -2\\0\\-1\\-2\\1\\0\\0 \end{bmatrix}$$

which can be used to create the linear combination

$$(-2)\mathbf{A}_1 + 0\mathbf{A}_2 + (-1)\mathbf{A}_3 + (-2)\mathbf{A}_4 + 1\mathbf{A}_5 + 0\mathbf{A}_6 + 0\mathbf{A}_7 = \mathbf{0}$$

This can then be arranged and solved for A_5 , resulting in A_5 expressed as a linear combination of $\{A_1, A_3, A_4\}$,

$$\mathbf{A}_5 = 2\mathbf{A}_1 + 1\mathbf{A}_3 + 2\mathbf{A}_4$$

This means that \mathbf{A}_5 is surplus, and we can create W just as well with a smaller set with this vector removed,

$$W = \langle \{\mathbf{A}_1, \mathbf{A}_3, \mathbf{A}_4, \mathbf{A}_6, \mathbf{A}_7\} \rangle$$

Do it again, set the free variable x_6 to one, and set the other free variables to zero. Then

a solution to $\mathcal{LS}(B, \mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} -1\\0\\3\\6\\0\\1\\0 \end{bmatrix}$$

which can be used to create the linear combination

$$(-1)\mathbf{A}_1 + 0\mathbf{A}_2 + 3\mathbf{A}_3 + 6\mathbf{A}_4 + 0\mathbf{A}_5 + 1\mathbf{A}_6 + 0\mathbf{A}_7 = \mathbf{0}$$

This can then be arranged and solved for A_6 , resulting in A_6 expressed as a linear combination of $\{A_1, A_3, A_4\}$,

$$\mathbf{A}_6 = 1\mathbf{A}_1 + (-3)\mathbf{A}_3 + (-6)\mathbf{A}_4$$

This means that \mathbf{A}_6 is surplus, and we can create W just as well with a smaller set with this vector removed,

$$W = \langle \{\mathbf{A}_1, \, \mathbf{A}_3, \, \mathbf{A}_4, \, \mathbf{A}_7\} \rangle$$

Set the free variable x_7 to one, and set the other free variables to zero. Then a solution

to $\mathcal{LS}(B, \mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} 3\\0\\-5\\-6\\0\\0\\1 \end{bmatrix}$$

which can be used to create the linear combination

$$3\mathbf{A}_1 + 0\mathbf{A}_2 + (-5)\mathbf{A}_3 + (-6)\mathbf{A}_4 + 0\mathbf{A}_5 + 0\mathbf{A}_6 + 1\mathbf{A}_7 = \mathbf{0}$$

This can then be arranged and solved for A_7 , resulting in A_7 expressed as a linear combination of $\{A_1, A_3, A_4\}$,

$$\mathbf{A}_7 = (-3)\mathbf{A}_1 + 5\mathbf{A}_3 + 6\mathbf{A}_4$$

This means that A_7 is surplus, and we can create W just as well with a smaller set with this vector removed,

$$W = \langle \{\mathbf{A}_1, \, \mathbf{A}_3, \, \mathbf{A}_4\} \rangle$$

You might think we could keep this up, but we have run out of free variables. And not coincidentally, the set $\{A_1, A_3, A_4\}$ is linearly independent (check this!). It should be clear how each free variable was used to eliminate the corresponding column from the set used to span the column space, as this will be the essence of the proof of the next theorem. The

column vectors in S were not chosen entirely at random, they are the columns of Archetype I [1714]. See if you can mimic this example using the columns of Archetype J [1723]. Go ahead, we'll go grab a cup of coffee and be back before you finish up.

For extra credit, notice that the vector

$$\mathbf{b} = \begin{bmatrix} 3\\9\\1\\4 \end{bmatrix}$$

is the vector of constants in the definition of Archetype I [1714]. Since the system $\mathcal{LS}(A, \mathbf{b})$ is consistent, we know by Theorem SLSLC [229] that \mathbf{b} is a linear combination of the columns of A, or stated equivalently, $\mathbf{b} \in W$. This means that \mathbf{b} must also be a linear combination of just the three columns $\mathbf{A}_1, \mathbf{A}_3, \mathbf{A}_4$. Can you find such a linear combination? Did you notice that there is just a single (unique) answer? Hmmmn.

Example COV [370] deserves your careful attention, since this important example motivates the following very fundamental theorem.

Theorem BS

Basis of a Span

Suppose that $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n}$ is a set of column vectors. Define $W = \langle S \rangle$ and let A be the matrix whose columns are the vectors from S. Let B be the reduced row-echelon form of A, with $D = {d_1, d_2, d_3, \dots, d_r}$ the set of column indices corresponding to the pivot columns of B. Then

1. $T = {\mathbf{v}_{d_1}, \mathbf{v}_{d_2}, \mathbf{v}_{d_3}, \dots, \mathbf{v}_{d_r}}$ is a linearly independent set.

2.
$$W = \langle T \rangle$$
.

Proof To prove that T is linearly independent, begin with a relation of linear dependence on T,

$$\mathbf{0} = \alpha_1 \mathbf{v}_{d_1} + \alpha_2 \mathbf{v}_{d_2} + \alpha_3 \mathbf{v}_{d_3} + \ldots + \alpha_r \mathbf{v}_{d_r}$$

and we will try to conclude that the only possibility for the scalars α_i is that they are all zero. Denote the non-pivot columns of B by $F = \{f_1, f_2, f_3, \ldots, f_{n-r}\}$. Then we can preserve the equality by adding a big fat zero to the linear combination,

$$\mathbf{0} = \alpha_1 \mathbf{v}_{d_1} + \alpha_2 \mathbf{v}_{d_2} + \alpha_3 \mathbf{v}_{d_3} + \ldots + \alpha_r \mathbf{v}_{d_r} + 0 \mathbf{v}_{f_1} + 0 \mathbf{v}_{f_2} + 0 \mathbf{v}_{f_3} + \ldots + 0 \mathbf{v}_{f_{n-r}}$$

By Theorem SLSLC [229], the scalars in this linear combination (suitably reordered) are a solution to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. But notice that this is the solution obtained by setting each free variable to zero. If we consider the description of a solution vector in the conclusion of Theorem VFSLS [244], in the case of a homogeneous system, then we see that if all the free variables are set to zero the resulting solution vector is trivial (all zeros). So it must be that $\alpha_i = 0, 1 \leq i \leq r$. This implies by Definition LICV [317] that T is a linearly independent set.

The second conclusion of this theorem is an equality of sets (Definition SE [1604]). Since T is a subset of S, any linear combination of elements of the set T can also be viewed as a

 \square

linear combination of elements of the set S. So $\langle T \rangle \subseteq \langle S \rangle = W$. It remains to prove that $W = \langle S \rangle \subseteq \langle T \rangle$.

For each k, $1 \le k \le n-r$, form a solution **x** to $\mathcal{LS}(A, \mathbf{0})$ by setting the free variables as follows:

$$x_{f_1} = 0$$
 $x_{f_2} = 0$ $x_{f_3} = 0$... $x_{f_k} = 1$... $x_{f_{n-r}} = 0$

By Theorem VFSLS [244], the remainder of this solution vector is given by,

$$x_{d_1} = -[B]_{1,f_k}$$
 $x_{d_2} = -[B]_{2,f_k}$ $x_{d_3} = -[B]_{3,f_k}$ \dots $x_{d_r} = -[B]_{r,f_k}$

From this solution, we obtain a relation of linear dependence on the columns of A,

$$-[B]_{1,f_k}\mathbf{v}_{d_1}-[B]_{2,f_k}\mathbf{v}_{d_2}-[B]_{3,f_k}\mathbf{v}_{d_3}-\ldots-[B]_{r,f_k}\mathbf{v}_{d_r}+1\mathbf{v}_{f_k}=\mathbf{0}$$

which can be arranged as the equality

$$\mathbf{v}_{f_k} = [B]_{1,f_k} \, \mathbf{v}_{d_1} + [B]_{2,f_k} \, \mathbf{v}_{d_2} + [B]_{3,f_k} \, \mathbf{v}_{d_3} + \ldots + [B]_{r,f_k} \, \mathbf{v}_{d_r}$$

Now, suppose we take an arbitrary element, \mathbf{w} , of $W = \langle S \rangle$ and write it as a linear combination of the elements of S, but with the terms organized according to the indices in D and F,

$$\mathbf{w} = \alpha_1 \mathbf{v}_{d_1} + \alpha_2 \mathbf{v}_{d_2} + \alpha_3 \mathbf{v}_{d_3} + \ldots + \alpha_r \mathbf{v}_{d_r} + \beta_1 \mathbf{v}_{f_1} + \beta_2 \mathbf{v}_{f_2} + \beta_3 \mathbf{v}_{f_3} + \ldots + \beta_{n-r} \mathbf{v}_{f_{n-r}}$$

From the above, we can replace each \mathbf{v}_{f_j} by a linear combination of the \mathbf{v}_{d_i} ,

$$\mathbf{w} = \alpha_{1}\mathbf{v}_{d_{1}} + \alpha_{2}\mathbf{v}_{d_{2}} + \alpha_{3}\mathbf{v}_{d_{3}} + \dots + \alpha_{r}\mathbf{v}_{d_{r}} + \beta_{1}\left([B]_{1,f_{1}}\mathbf{v}_{d_{1}} + [B]_{2,f_{1}}\mathbf{v}_{d_{2}} + [B]_{3,f_{1}}\mathbf{v}_{d_{3}} + \dots + [B]_{r,f_{1}}\mathbf{v}_{d_{r}}\right) + \beta_{2}\left([B]_{1,f_{2}}\mathbf{v}_{d_{1}} + [B]_{2,f_{2}}\mathbf{v}_{d_{2}} + [B]_{3,f_{2}}\mathbf{v}_{d_{3}} + \dots + [B]_{r,f_{2}}\mathbf{v}_{d_{r}}\right) + \beta_{3}\left([B]_{1,f_{3}}\mathbf{v}_{d_{1}} + [B]_{2,f_{3}}\mathbf{v}_{d_{2}} + [B]_{3,f_{3}}\mathbf{v}_{d_{3}} + \dots + [B]_{r,f_{3}}\mathbf{v}_{d_{r}}\right) + \vdots \\ \beta_{n-r}\left([B]_{1,f_{n-r}}\mathbf{v}_{d_{1}} + [B]_{2,f_{n-r}}\mathbf{v}_{d_{2}} + [B]_{3,f_{n-r}}\mathbf{v}_{d_{3}} + \dots + [B]_{r,f_{n-r}}\mathbf{v}_{d_{r}}\right)$$

With repeated applications of several of the properties of Theorem VSPCV [205] we can rearrange this expression as,

$$= \left(\alpha_{1} + \beta_{1} \left[B\right]_{1,f_{1}} + \beta_{2} \left[B\right]_{1,f_{2}} + \beta_{3} \left[B\right]_{1,f_{3}} + \dots + \beta_{n-r} \left[B\right]_{1,f_{n-r}}\right) \mathbf{v}_{d_{1}} + \left(\alpha_{2} + \beta_{1} \left[B\right]_{2,f_{1}} + \beta_{2} \left[B\right]_{2,f_{2}} + \beta_{3} \left[B\right]_{2,f_{3}} + \dots + \beta_{n-r} \left[B\right]_{2,f_{n-r}}\right) \mathbf{v}_{d_{2}} + \left(\alpha_{3} + \beta_{1} \left[B\right]_{3,f_{1}} + \beta_{2} \left[B\right]_{3,f_{2}} + \beta_{3} \left[B\right]_{3,f_{3}} + \dots + \beta_{n-r} \left[B\right]_{3,f_{n-r}}\right) \mathbf{v}_{d_{3}} + \\ \vdots \\ \left(\alpha_{r} + \beta_{1} \left[B\right]_{r,f_{1}} + \beta_{2} \left[B\right]_{r,f_{2}} + \beta_{3} \left[B\right]_{r,f_{3}} + \dots + \beta_{n-r} \left[B\right]_{r,f_{n-r}}\right) \mathbf{v}_{d_{r}}$$

This mess expresses the vector \mathbf{w} as a linear combination of the vectors in

$$T = \{\mathbf{v}_{d_1}, \, \mathbf{v}_{d_2}, \, \mathbf{v}_{d_3}, \, \dots \, \mathbf{v}_{d_r}\}$$

thus saying that $\mathbf{w} \in \langle T \rangle$. Therefore, $W = \langle S \rangle \subseteq \langle T \rangle$.

In Example COV [370], we tossed-out vectors one at a time. But in each instance, we rewrote the offending vector as a linear combination of those vectors that corresponded to the pivot columns of the reduced row-echelon form of the matrix of columns. In the proof of Theorem BS [376], we accomplish this reduction in one big step. In Example COV [370] we arrived at a linearly independent set at exactly the same moment that we ran out of free variables to exploit. This was not a coincidence, it is the substance of our conclusion of linear independence in Theorem BS [376].

Here's a straightforward application of Theorem BS [376].

Example RSC4

Reducing a span in \mathbb{C}^4

Begin with a set of five vectors from \mathbb{C}^4 ,

$$S = \left\{ \begin{bmatrix} 1\\1\\2\\1 \end{bmatrix}, \begin{bmatrix} 2\\2\\4\\2 \end{bmatrix}, \begin{bmatrix} 2\\0\\-1\\1 \end{bmatrix}, \begin{bmatrix} 7\\1\\-1\\4 \end{bmatrix}, \begin{bmatrix} 0\\2\\5\\1 \end{bmatrix} \right\}$$

and let $W = \langle S \rangle$. To arrive at a (smaller) linearly independent set, follow the procedure described in Theorem BS [376]. Place the vectors from S into a matrix as columns, and

row-reduce,

Columns 1 and 3 are the pivot columns $(D = \{1, 3\})$ so the set

$$T = \left\{ \begin{bmatrix} 1\\1\\2\\1 \end{bmatrix}, \begin{bmatrix} 2\\0\\-1\\1 \end{bmatrix} \right\}$$

is linearly independent and $\langle T \rangle = \langle S \rangle = W$. Boom!

Since the reduced row-echelon form of a matrix is unique (Theorem RREFU [70]), the procedure of Theorem BS [376] leads us to a unique set T. However, there is a wide variety of possibilities for sets T that are linearly independent and which can be employed in a span to create W. Without proof, we list two other possibilities:

$$T' = \left\{ \begin{bmatrix} 2\\2\\4\\2 \end{bmatrix}, \begin{bmatrix} 2\\0\\-1\\1 \end{bmatrix} \right\}$$

$$T^* = \left\{ \begin{bmatrix} 3\\1\\1\\2 \end{bmatrix}, \begin{bmatrix} -1\\1\\3\\0 \end{bmatrix} \right\}$$

Can you prove that T' and T^* are linearly independent sets and $W = \langle S \rangle = \langle T' \rangle = \langle T^* \rangle$?

Example RES Reworking elements of a span

Begin with a set of five vectors from \mathbb{C}^4 ,

$$R = \left\{ \begin{bmatrix} 2\\1\\3\\2 \end{bmatrix}, \begin{bmatrix} -1\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} -8\\-1\\-9\\-4 \end{bmatrix}, \begin{bmatrix} 3\\1\\-1\\-1\\-2 \end{bmatrix}, \begin{bmatrix} -10\\-1\\-1\\-1\\4 \end{bmatrix} \right\}$$

It is easy to create elements of $X = \langle R \rangle$ — we will create one at random,

$$\mathbf{y} = 6 \begin{bmatrix} 2\\1\\3\\2 \end{bmatrix} + (-7) \begin{bmatrix} -1\\1\\0\\1 \end{bmatrix} + 1 \begin{bmatrix} -8\\-1\\-9\\-4 \end{bmatrix} + 6 \begin{bmatrix} 3\\1\\-1\\-2 \end{bmatrix} + 2 \begin{bmatrix} -10\\-1\\-1\\-1\\4 \end{bmatrix} = \begin{bmatrix} 9\\2\\1\\-3 \end{bmatrix}$$

We know we can replace R by a smaller set (since it is obviously linearly dependent by

Theorem MVSLD [328]) that will create the same span. Here goes,

$$\begin{bmatrix} 2 & -1 & -8 & 3 & -10 \\ 1 & 1 & -1 & 1 & -1 \\ 3 & 0 & -9 & -1 & -1 \\ 2 & 1 & -4 & -2 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -3 & 0 & -1 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So, if we collect the first, second and fourth vectors from R,

$$P = \left\{ \begin{bmatrix} 2\\1\\3\\2 \end{bmatrix}, \begin{bmatrix} -1\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} 3\\1\\-1\\-2 \end{bmatrix} \right\}$$

then P is linearly independent and $\langle P \rangle = \langle R \rangle = X$ by Theorem BS [376]. Since we built **y** as an element of $\langle R \rangle$ it must also be an element of $\langle P \rangle$. Can we write **y** as a linear combination of just the three vectors in P? The answer is, of course, yes. But let's compute an explicit linear combination just for fun. By Theorem SLSLC [229] we can get such a linear combination by solving a system of equations with the column vectors of R as the columns of a coefficient matrix, and **y** as the vector of constants. Employing an augmented matrix to solve this system,

$$\begin{bmatrix} 2 & -1 & 3 & 9 \\ 1 & 1 & 1 & 2 \\ 3 & 0 & -1 & 1 \\ 2 & 1 & -2 & -3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So we see, as expected, that

$$1\begin{bmatrix} 2\\1\\3\\2 \end{bmatrix} + (-1)\begin{bmatrix} -1\\1\\0\\1 \end{bmatrix} + 2\begin{bmatrix} 3\\1\\-1\\-2 \end{bmatrix} = \begin{bmatrix} 9\\2\\1\\-3 \end{bmatrix} = \mathbf{y}$$

A key feature of this example is that the linear combination that expresses \mathbf{y} as a linear combination of the vectors in P is unique. This is a consequence of the linear independence of P. The linearly independent set P is smaller than R, but still just (barely) big enough to create elements of the set $X = \langle R \rangle$. There are many, many ways to write \mathbf{y} as a linear combination of the five vectors in R (the appropriate system of equations to verify this claim has two free variables in the description of the solution set), yet there is precisely one way to write \mathbf{y} as a linear combination of the three vectors in P.

Subsection READ Reading Questions

1. Let S be the linearly dependent set of three vectors below.

$$S = \left\{ \begin{bmatrix} 1\\10\\100\\1000 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} 5\\23\\203\\2003 \end{bmatrix} \right\}$$

Write one vector from S as a linear combination of the other two and include this vector equality in your response. (You should be able to do this on sight, rather than doing some computations.) Convert this expression into a nontrivial relation of linear dependence on S.

- 2. Explain why the word "dependent" is used in the definition of linear dependence.
- 3. Suppose that $Y = \langle P \rangle = \langle Q \rangle$, where P is a linearly dependent set and Q is linearly independent. Would you rather use P or Q to describe Y? Why?

Subsection EXC Exercises

C20 Let T be the set of columns of the matrix B below. Define $W = \langle T \rangle$. Find a set R so that (1) R has 3 vectors, (2) R is a subset of T, and (3) $W = \langle R \rangle$.

$$B = \begin{bmatrix} -3 & 1 & -2 & 7\\ -1 & 2 & 1 & 4\\ 1 & 1 & 2 & -1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [389]

C40 Verify that the set $R' = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4}$ at the end of Example RSC5 [366] is linearly independent.

Contributed by Robert Beezer

C50 Consider the set of vectors from \mathbb{C}^3 , W, given below. Find a linearly independent set T that contains three vectors from W and such that $\langle W \rangle = \langle T \rangle$.

$$W = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4, \, \mathbf{v}_5\} = \left\{ \begin{bmatrix} 2\\1\\1 \end{bmatrix}, \, \begin{bmatrix} -1\\-1\\1 \end{bmatrix}, \, \begin{bmatrix} 1\\2\\3 \end{bmatrix}, \, \begin{bmatrix} 3\\1\\3 \end{bmatrix}, \, \begin{bmatrix} 0\\1\\-3 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [389]

C51 Given the set S below, find a linearly independent set T so that $\langle T \rangle = \langle S \rangle$.

$$S = \left\{ \begin{bmatrix} 2\\-1\\2 \end{bmatrix}, \begin{bmatrix} 3\\0\\1 \end{bmatrix}, \begin{bmatrix} 1\\1\\-1 \end{bmatrix}, \begin{bmatrix} 5\\-1\\3 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [390]

C52 Let W be the span of the set of vectors S below, $W = \langle S \rangle$. Find a set T so that 1) the span of T is W, $\langle T \rangle = W$, (2) T is a linearly independent set, and (3) T is a subset of S.

$$S = \left\{ \begin{bmatrix} 1\\2\\-1 \end{bmatrix}, \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 4\\1\\-1 \end{bmatrix}, \begin{bmatrix} 3\\1\\1 \end{bmatrix}, \begin{bmatrix} 3\\-1\\0 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [391]

C55 Let *T* be the set of vectors
$$T = \left\{ \begin{bmatrix} 1\\ -1\\ 2 \end{bmatrix}, \begin{bmatrix} 3\\ 0\\ 1 \end{bmatrix}, \begin{bmatrix} 4\\ 2\\ 3 \end{bmatrix}, \begin{bmatrix} 3\\ 0\\ 6 \end{bmatrix} \right\}$$
. Find two different subsets of *T* named *R* and *S* are that *R* and *S* each contain three vectors, and so that

subsets of T, named R and S, so that R and S each contain three vectors, and so that $\langle R \rangle = \langle T \rangle$ and $\langle S \rangle = \langle T \rangle$. Prove that both R and S are linearly independent. Contributed by Robert Beezer Solution [392]

C70 Reprise Example RES [382] by creating a new version of the vector \mathbf{y} . In other words, form a new, different linear combination of the vectors in R to create a new vector \mathbf{y} (but

do not simplify the problem too much by choosing any of the five new scalars to be zero). Then express this new \mathbf{y} as a combination of the vectors in P.

Contributed by Robert Beezer

M10 At the conclusion of Example RSC4 [380] two alternative solutions, sets T' and T^* , are proposed. Verify these claims by proving that $\langle T \rangle = \langle T' \rangle$ and $\langle T \rangle = \langle T^* \rangle$. Contributed by Robert Beezer

T40 Suppose that \mathbf{v}_1 and \mathbf{v}_2 are any two vectors from \mathbb{C}^m . Prove the following set equality.

$$\langle \{\mathbf{v}_1, \, \mathbf{v}_2\} \rangle = \langle \{\mathbf{v}_1 + \mathbf{v}_2, \, \mathbf{v}_1 - \mathbf{v}_2\} \rangle$$

Contributed by Robert Beezer Solution [393]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [386]
Let
$$T = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\}$$
. The vector $\begin{bmatrix} 2\\ -1\\ 0\\ 1 \end{bmatrix}$ is a solution to the homogeneous system with the matrix B as the coefficient matrix (check this!). By Theorem SLSLC [229] it provides

the scalars for a linear combination of the columns of B (the vectors in T) that equals the zero vector, a relation of linear dependence on T,

$$2\mathbf{w}_1 + (-1)\mathbf{w}_2 + (1)\mathbf{w}_4 = \mathbf{0}$$

We can rearrange this equation by solving for \mathbf{w}_4 ,

$$\mathbf{w}_4 = (-2)\mathbf{w}_1 + \mathbf{w}_2$$

This equation tells us that the vector \mathbf{w}_4 is superfluous in the span construction that creates W. So $W = \langle \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} \rangle$. The requested set is $R = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

C50 Contributed by Robert Beezer Statement [386] To apply Theorem BS [376], we formulate a matrix A whose columns are \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 , \mathbf{v}_4 , \mathbf{v}_5 . Then we row-reduce A. After row-reducing, we obtain

$$\begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & 1 & -2 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

From this we see that the pivot columns are $D = \{1, 2, 3\}$. Thus

$$T = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3\} = \left\{ \begin{bmatrix} 2\\1\\1 \end{bmatrix}, \begin{bmatrix} -1\\-1\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\3 \end{bmatrix} \right\}$$

is a linearly independent set and $\langle T \rangle = W$. Compare this problem with Exercise LI.M50 [345].

C51 Contributed by Robert Beezer Statement [387] Theorem BS [376] says we can make a matrix with these four vectors as columns, row-reduce, and just keep the columns with indices in the set D. Here we go, forming the relevant matrix

and row-reducing,

$$\begin{bmatrix} 2 & 3 & 1 & 5 \\ -1 & 0 & 1 & -1 \\ 2 & 1 & -1 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & -1 & 1 \\ 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Analyzing the row-reduced version of this matrix, we see that the first two columns are pivot columns, so $D = \{1, 2\}$. Theorem BS [376] says we need only "keep" the first two columns

to create a set with the requisite properties,

$$T = \left\{ \begin{bmatrix} 2\\-1\\2 \end{bmatrix}, \begin{bmatrix} 3\\0\\1 \end{bmatrix} \right\}$$

C52 Contributed by Robert Beezer Statement [387]

This is a straight setup for the conclusion of Theorem BS [376]. The hypotheses of this theorem tell us to pack the vectors of W into the columns of a matrix and row-reduce,

$$\begin{bmatrix} 1 & 2 & 4 & 3 & 3 \\ 2 & -3 & 1 & 1 & -1 \\ -1 & 1 & -1 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Pivot columns have indices $D = \{1, 2, 4\}$. Theorem BS [376] tells us to form T with columns 1, 2 and 4 of S,

$$S = \left\{ \begin{bmatrix} 1\\2\\-1 \end{bmatrix}, \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 3\\1\\1 \end{bmatrix} \right\}$$

C55 Contributed by Robert Beezer Statement [387]

Let A be the matrix whose columns are the vectors in T. Then row-reduce A,

$$A \xrightarrow{\text{RREF}} B = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

From Theorem BS [376] we can form R by choosing the columns of A that correspond to the pivot columns of B. Theorem BS [376] also guarantees that R will be linearly independent.

$$R = \left\{ \begin{bmatrix} 1\\-1\\2 \end{bmatrix}, \begin{bmatrix} 3\\0\\1 \end{bmatrix}, \begin{bmatrix} 4\\2\\3 \end{bmatrix} \right\}$$

That was easy. To find S will require a bit more work. From B we can obtain a solution to $\mathcal{LS}(A, \mathbf{0})$, which by Theorem SLSLC [229] will provide a nontrivial relation of linear dependence on the columns of A, which are the vectors in T. To wit, choose the free variable x_4 to be 1, then $x_1 = -2$, $x_2 = 1$, $x_3 = -1$, and so

$$(-2)\begin{bmatrix}1\\-1\\2\end{bmatrix} + (1)\begin{bmatrix}3\\0\\1\end{bmatrix} + (-1)\begin{bmatrix}4\\2\\3\end{bmatrix} + (1)\begin{bmatrix}3\\0\\6\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

this equation can be rewritten with the second vector staying put, and the other three moving to the other side of the equality,

$$\begin{bmatrix} 3\\0\\1 \end{bmatrix} = (2) \begin{bmatrix} 1\\-1\\2 \end{bmatrix} + (1) \begin{bmatrix} 4\\2\\3 \end{bmatrix} + (-1) \begin{bmatrix} 3\\0\\6 \end{bmatrix}$$

We could have chosen other vectors to stay put, but may have then needed to divide by a nonzero scalar. This equation is enough to conclude that the second vector in T is "surplus" and can be replaced (see the careful argument in Example RSC5 [366]). So set

$$S = \left\{ \begin{bmatrix} 1\\-1\\2 \end{bmatrix}, \begin{bmatrix} 4\\2\\3 \end{bmatrix}, \begin{bmatrix} 3\\0\\6 \end{bmatrix} \right\}$$

and then $\langle S \rangle = \langle T \rangle$. T is also a linearly independent set, which we can show directly. Make a matrix C whose columns are the vectors in S. Row-reduce B and you will obtain the identity matrix I_3 . By Theorem LIVRN [325], the set S is linearly independent.

T40 Contributed by Robert Beezer Statement [388] This is an equality of sets, so Definition SE [1604] applies.

The "easy" half first. Show that $X = \langle \{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 - \mathbf{v}_2\} \rangle \subseteq \langle \{\mathbf{v}_1, \mathbf{v}_2\} \rangle = Y$. Choose $\mathbf{x} \in X$. Then $\mathbf{x} = a_1(\mathbf{v}_1 + \mathbf{v}_2) + a_2(\mathbf{v}_1 - \mathbf{v}_2)$ for some scalars a_1 and a_2 . Then,

$$\mathbf{x} = a_1(\mathbf{v}_1 + \mathbf{v}_2) + a_2(\mathbf{v}_1 - \mathbf{v}_2)$$

= $a_1\mathbf{v}_1 + a_1\mathbf{v}_2 + a_2\mathbf{v}_1 + (-a_2)\mathbf{v}_2$
= $(a_1 + a_2)\mathbf{v}_1 + (a_1 - a_2)\mathbf{v}_2$

which qualifies \mathbf{x} for membership in Y, as it is a linear combination of \mathbf{v}_1 , \mathbf{v}_2 .

Now show the opposite inclusion, $Y = \langle \{\mathbf{v}_1, \mathbf{v}_2\} \rangle \subseteq \langle \{\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 - \mathbf{v}_2\} \rangle = X$. Choose $\mathbf{y} \in Y$. Then there are scalars b_1, b_2 such that $\mathbf{y} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2$. Rearranging, we obtain,

$$\begin{aligned} \mathbf{y} &= b_1 \mathbf{v}_1 + b_2 \mathbf{v}_2 \\ &= \frac{b_1}{2} \left[(\mathbf{v}_1 + \mathbf{v}_2) + (\mathbf{v}_1 - \mathbf{v}_2) \right] + \frac{b_2}{2} \left[(\mathbf{v}_1 + \mathbf{v}_2) - (\mathbf{v}_1 - \mathbf{v}_2) \right] \\ &= \frac{b_1 + b_2}{2} \left(\mathbf{v}_1 + \mathbf{v}_2 \right) + \frac{b_1 - b_2}{2} \left(\mathbf{v}_1 - \mathbf{v}_2 \right) \end{aligned}$$

This is an expression for \mathbf{y} as a linear combination of $\mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{v}_1 - \mathbf{v}_2$, earning \mathbf{y} membership in X. Since X is a subset of Y, and vice versa, we see that X = Y, as desired.

Section O Orthogonality

In this section we define a couple more operations with vectors, and prove a few theorems. At first blush these definitions and results will not appear central to what follows, but we will make use of them at key points in the remainder of the course (such as Section MINM [537], Section OD [1414]). Because we have chosen to use \mathbb{C} as our set of scalars, this subsection is a bit more, uh, ... complex than it would be for the real numbers. We'll explain as we go along how things get easier for the real numbers \mathbb{R} . If you haven't already, now would be a good time to review some of the basic properties of arithmetic with complex numbers described in Section CNO [1594]. With that done, we can extend the basics of complex number arithmetic to our study of vectors in \mathbb{C}^m .

Subsection CAV Complex Arithmetic and Vectors

We know how the addition and multiplication of complex numbers is employed in defining the operations for vectors in \mathbb{C}^m (Definition CVA [201] and Definition CVSM [203]). We can

also extend the idea of the conjugate to vectors.

Definition CCCV Complex Conjugate of a Column Vector

Suppose that **u** is a vector from \mathbb{C}^m . Then the conjugate of the vector, $\overline{\mathbf{u}}$, is defined by

$$\left[\overline{\mathbf{u}}\right]_i = \overline{\left[\mathbf{u}\right]_i} \qquad \qquad 1 \le i \le m$$

(This definition contains Notation CCCV.)

With this definition we can show that the conjugate of a column vector behaves as we would expect with regard to vector addition and scalar multiplication.

Theorem CRVA Conjugation Respects Vector Addition Suppose \mathbf{x} and \mathbf{y} are two vectors from \mathbb{C}^m . Then

$$\overline{\mathbf{x} + \mathbf{y}} = \overline{\mathbf{x}} + \overline{\mathbf{y}}$$

Proof For each $1 \le i \le m$,

$$\overline{[\mathbf{x} + \mathbf{y}]}_i = \overline{[\mathbf{x} + \mathbf{y}]}_i$$

= $\overline{[\mathbf{x}]_i + [\mathbf{y}]_i}$

Definition CCCV [396] Definition CVA [201] \wedge

$$= [\mathbf{x}]_i + [\mathbf{y}]_i$$
$$= [\overline{\mathbf{x}}]_i + [\overline{\mathbf{y}}]_i$$
$$= [\overline{\mathbf{x}} + \overline{\mathbf{y}}]_i$$

Theorem CCRA [1599] Definition CCCV [396] Definition CVA [201]

Then by Definition CVE [199] we have $\overline{\mathbf{x} + \mathbf{y}} = \overline{\mathbf{x}} + \overline{\mathbf{y}}$.

Theorem CRSM Conjugation Respects Vector Scalar Multiplication Suppose \mathbf{x} is a vector from \mathbb{C}^m , and $\alpha \in \mathbb{C}$ is a scalar. Then

 $\overline{\alpha}\overline{\mathbf{x}} = \overline{\alpha}\,\overline{\mathbf{x}}$

Proof For $1 \le i \le m$,

$$\overline{\alpha \mathbf{x}}]_i = \overline{[\alpha \mathbf{x}]_i} = \overline{\alpha [\mathbf{x}]_i} = \overline{\alpha [\mathbf{x}]_i} = \overline{\alpha} \overline{[\mathbf{x}]_i} = \overline{\alpha} \overline{[\mathbf{x}]_i} = \overline{\alpha} \overline{[\mathbf{x}]_i} = \overline{\alpha} \overline{[\mathbf{x}]_i} = \overline{\alpha} \overline{[\mathbf{x}]_i}$$

Definition CCCV [396] Definition CVSM [203] Theorem CCRM [1599] Definition CCCV [396] Definition CVSM [203] \square

Then by Definition CVE [199] we have $\overline{\alpha \mathbf{x}} = \overline{\alpha} \, \overline{\mathbf{x}}$.

These two theorems together tell us how we can "push" complex conjugation through linear combinations.

Subsection IP Inner products

Definition IP Inner Product

Given the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$ the inner product of \mathbf{u} and \mathbf{v} is the scalar quantity in \mathbb{C} ,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \overline{[\mathbf{u}]_1} [\mathbf{v}]_1 + \overline{[\mathbf{u}]_2} [\mathbf{v}]_2 + \overline{[\mathbf{u}]_3} [\mathbf{v}]_3 + \dots + \overline{[\mathbf{u}]_m} [\mathbf{v}]_m = \sum_{i=1}^m \overline{[\mathbf{u}]_i} [\mathbf{v}]_i$$

(This definition contains Notation IP.)

This operation is a bit different in that we begin with two vectors but produce a scalar. Computing one is straightforward.

Example CSIP Computing some inner products \wedge

The scalar product of

$$\mathbf{u} = \begin{bmatrix} 2+3i\\5+2i\\-3+i \end{bmatrix} \qquad \text{and} \qquad \mathbf{v} = \begin{bmatrix} 1+2i\\-4+5i\\0+5i \end{bmatrix}$$

is

$$\langle \mathbf{u}, \mathbf{v} \rangle = (\overline{2+3i})(1+2i) + (\overline{5+2i})(-4+5i) + (\overline{-3+i})(0+5i) = (2-3i)(1+2i) + (5-2i)(-4+5i) + (-3-i)(0+5i) = (8+i) + (-10+33i) + (5-15i) = 3+19i$$

The scalar product of

$$\mathbf{w} = \begin{bmatrix} 2\\4\\-3\\2\\8 \end{bmatrix} \qquad \text{and} \qquad \mathbf{x} = \begin{bmatrix} 3\\1\\0\\-1\\-2 \end{bmatrix}$$

is

 $\langle \mathbf{w}, \, \mathbf{x} \rangle = (\overline{2})3 + (\overline{4})1 + (\overline{-3})0 + (\overline{2})(-1) + (\overline{8})(-2) = 2(3) + 4(1) + (-3)0 + 2(-1) + 8(-2) = -8.$

Version 2.99

 \boxtimes

In the case where the entries of our vectors are all real numbers (as in the second part of Example CSIP [398]), the computation of the inner product may look familiar and be known to you as a **dot product** or **scalar product**. So you can view the inner product as a generalization of the scalar product to vectors from \mathbb{C}^m (rather than \mathbb{R}^m).

Note that we have chosen to conjugate the entries of the *first* vector listed in the inner product, while it is almost equally feasible to conjugate entries from the *second* vector instead. In particular, prior to Version 2.90, we did use the latter definition, and this has now changed to the former, with resulting adjustments propogated up through Section CB [1355] (only). However, conjugating the first vector leads to much nicer formulas for certain matrix decompositions and also shortens some proofs.

There are several quick theorems we can now prove, and they will each be useful later.

Theorem IPVA Inner Product and Vector Addition Suppose $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^m$. Then

> 1. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ 2. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

Proof The proofs of the two parts are very similar, with the second one requiring just a bit more effort due to the conjugation that occurs. We will prove part 1 and you can prove

part 2 (Exercise O.T10 [422]).

$$\begin{split} \langle \mathbf{u} + \mathbf{v}, \, \mathbf{w} \rangle &= \sum_{i=1}^{m} \overline{[\mathbf{u} + \mathbf{v}]_{i}} \, [\mathbf{w}]_{i} & \mathbf{I} \\ &= \sum_{i=1}^{m} \left(\overline{[\mathbf{u}]_{i}} + \overline{[\mathbf{v}]_{i}} \right) [\mathbf{w}]_{i} & \mathbf{I} \\ &= \sum_{i=1}^{m} \left(\overline{[\mathbf{u}]_{i}} + \overline{[\mathbf{v}]_{i}} \right) [\mathbf{w}]_{i} & \mathbf{I} \\ &= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} \, [\mathbf{w}]_{i} + \overline{[\mathbf{v}]_{i}} \, [\mathbf{w}]_{i} & \mathbf{I} \\ &= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} \, [\mathbf{w}]_{i} + \overline{[\mathbf{v}]_{i}} \, [\mathbf{w}]_{i} & \mathbf{I} \\ &= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} \, [\mathbf{w}]_{i} + \sum_{i=1}^{m} \overline{[\mathbf{v}]_{i}} \, [\mathbf{w}]_{i} & \mathbf{I} \\ &= \langle \mathbf{u}, \, \mathbf{w} \rangle + \langle \mathbf{v}, \, \mathbf{w} \rangle & \mathbf{I} \end{split}$$

Definition IP [398]

Theorem CCRA [1599]

Property DCN [1597]

Property CACN [1597]

Definition IP [398]

Theorem IPSM Inner Product and Scalar Multiplication Suppose $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$ and $\alpha \in \mathbb{C}$. Then

1.
$$\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \overline{\alpha} \langle \mathbf{u}, \mathbf{v} \rangle$$

2. $\langle \mathbf{u}, \alpha \mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$

 \square

Proof The proofs of the two parts are very similar, with the second one requiring just a bit more effort due to the conjugation that occurs. We will prove part 1 and you can prove part 2 (Exercise O.T11 [422]).

$$\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{m} \overline{[\alpha \mathbf{u}]_{i}} [\mathbf{v}]_{i} \qquad \text{Det}$$

$$= \sum_{i=1}^{m} \overline{\alpha} \overline{[\mathbf{u}]_{i}} [\mathbf{v}]_{i} \qquad \text{Det}$$

$$= \sum_{i=1}^{m} \overline{\alpha} \overline{[\mathbf{u}]_{i}} [\mathbf{v}]_{i} \qquad \text{Th}$$

$$= \overline{\alpha} \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} [\mathbf{v}]_{i} \qquad \text{Pr}$$

$$= \overline{\alpha} \langle \mathbf{u}, \mathbf{v} \rangle \qquad \text{Det}$$

Definition IP [398]

Definition CVSM [203]

Theorem CCRM [1599]

Property DCN [1597]

Definition IP [398]

Theorem IPAC Inner Product is Anti-Commutative

Suppose that **u** and **v** are vectors in \mathbb{C}^m . Then $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$.

Proof

$$\begin{aligned} \mathbf{u}, \, \mathbf{v} \rangle &= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} \, [\mathbf{v}]_{i} \\ &= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} \, \overline{\overline{[\mathbf{v}]_{i}}} \\ &= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i} \, \overline{[\mathbf{v}]_{i}}} \\ &= \overline{\left(\sum_{i=1}^{m} [\mathbf{u}]_{i} \, \overline{[\mathbf{v}]_{i}}\right)} \\ &= \overline{\left(\sum_{i=1}^{m} \overline{[\mathbf{v}]_{i}} \, [\mathbf{u}]_{i}\right)} \\ &= \overline{\left(\sum_{i=1}^{m} \overline{[\mathbf{v}]_{i}} \, [\mathbf{u}]_{i}\right)} \end{aligned}$$

Definition IP [398]

Theorem CCT [1600]

Theorem CCRM [1599]

Theorem CCRA [1599]

Property CMCN [1597]

Definition IP [398]

 \square

Subsection N Norm

If treating linear algebra in a more geometric fashion, the length of a vector occurs naturally, and is what you would expect from its name. With complex numbers, we will define a similar function. Recall that if c is a complex number, then |c| denotes its modulus (Definition MCN [1601]).

Definition NV Norm of a Vector The norm of the vector \mathbf{u} is the scalar quantity in \mathbb{C}

$$\|\mathbf{u}\| = \sqrt{|[\mathbf{u}]_1|^2 + |[\mathbf{u}]_2|^2 + |[\mathbf{u}]_3|^2 + \dots + |[\mathbf{u}]_m|^2} = \sqrt{\sum_{i=1}^m |[\mathbf{u}]_i|^2}$$

(This definition contains Notation NV.)

Computing a norm is also easy to do.

Example CNSV Computing the norm of some vectors The norm of

$$\mathbf{u} = \begin{bmatrix} 3+2i\\1-6i\\2+4i\\2+i \end{bmatrix}$$

is

$$\|\mathbf{u}\| = \sqrt{|3+2i|^2 + |1-6i|^2 + |2+4i|^2 + |2+i|^2} = \sqrt{13+37+20+5} = \sqrt{75} = 5\sqrt{3}.$$

The norm of

$$\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 2 \\ 4 \\ -3 \end{bmatrix}$$

is

$$\|\mathbf{v}\| = \sqrt{|3|^2 + |-1|^2 + |2|^2 + |4|^2 + |-3|^2} = \sqrt{3^2 + 1^2 + 2^2 + 4^2 + 3^2} = \sqrt{39}.$$

Notice how the norm of a vector with real number entries is just the length of the vector. Inner products and norms are related by the following theorem.

 \boxtimes

Theorem IPN Inner Products and Norms

Suppose that \mathbf{u} is a vector in \mathbb{C}^m . Then $\|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u} \rangle$.

 \mathbf{Proof}

$$\|\mathbf{u}\|^{2} = \left(\sqrt{\sum_{i=1}^{m} |[\mathbf{u}]_{i}|^{2}}\right)^{2}$$
 Definition NV [404]
$$= \sum_{i=1}^{m} |[\mathbf{u}]_{i}|^{2}$$

$$= \sum_{i=1}^{m} \overline{[\mathbf{u}]_{i}} [\mathbf{u}]_{i}$$
 Definition MCN [1601]
$$= \langle \mathbf{u}, \mathbf{u} \rangle$$
 Definition IP [398]

When our vectors have entries only from the real numbers Theorem IPN [406] says that the dot product of a vector with itself is equal to the length of the vector squared.

Theorem PIP Positive Inner Products

Suppose that **u** is a vector in \mathbb{C}^m . Then $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$. \Box **Proof** From the proof of Theorem IPN [406] we see that

$$\langle \mathbf{u}, \, \mathbf{u} \rangle = |[\mathbf{u}]_1|^2 + |[\mathbf{u}]_2|^2 + |[\mathbf{u}]_3|^2 + \dots + |[\mathbf{u}]_m|^2$$

Since each modulus is squared, every term is positive, and the sum must also be positive. (Notice that in general the inner product is a complex number and cannot be compared with zero, but in the special case of $\langle \mathbf{u}, \mathbf{u} \rangle$ the result is a real number.) The phrase, "with equality if and only if" means that we want to show that the statement $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ (i.e. with equality) is equivalent ("if and only if") to the statement $\mathbf{u} = \mathbf{0}$.

If $\mathbf{u} = \mathbf{0}$, then it is a straightforward computation to see that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. In the other direction, assume that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. As before, $\langle \mathbf{u}, \mathbf{u} \rangle$ is a sum of moduli. So we have

$$0 = \langle \mathbf{u}, \, \mathbf{u} \rangle = |[\mathbf{u}]_1|^2 + |[\mathbf{u}]_2|^2 + |[\mathbf{u}]_3|^2 + \dots + |[\mathbf{u}]_m|^2$$

Now we have a sum of squares equaling zero, so each term must be zero. Then by similar logic, $|[\mathbf{u}]_i| = 0$ will imply that $[\mathbf{u}]_i = 0$, since 0 + 0i is the only complex number with zero modulus. Thus every entry of \mathbf{u} is zero and so $\mathbf{u} = \mathbf{0}$, as desired.

Notice that Theorem PIP [406] contains three implications:

$$\mathbf{u} \in \mathbb{C}^m \Rightarrow \langle \mathbf{u}, \, \mathbf{u} \rangle \ge 0$$
$$\mathbf{u} = \mathbf{0} \Rightarrow \langle \mathbf{u}, \, \mathbf{u} \rangle = 0$$
$$\langle \mathbf{u}, \, \mathbf{u} \rangle = 0 \Rightarrow \mathbf{u} = \mathbf{0}$$

The results contained in Theorem PIP [406] are summarized by saying "the inner product is **positive definite**."

Subsection OV Orthogonal Vectors

"Orthogonal" is a generalization of "perpendicular." You may have used mutually perpendicular vectors in a physics class, or you may recall from a calculus class that perpendicular vectors have a zero dot product. We will now extend these ideas into the realm of higher dimensions and complex scalars.

Definition OV

Orthogonal Vectors

A pair of vectors, **u** and **v**, from \mathbb{C}^m are **orthogonal** if their inner product is zero, that is, $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Example TOV Two orthogonal vectors

The vectors

$$\mathbf{u} = \begin{bmatrix} 2+3i\\4-2i\\1+i\\1+i \end{bmatrix} \qquad \qquad \mathbf{v} = \begin{bmatrix} 1-i\\2+3i\\4-6i\\1 \end{bmatrix}$$

are orthogonal since

$$\langle \mathbf{u}, \mathbf{v} \rangle = (2 - 3i)(1 - i) + (4 + 2i)(2 + 3i) + (1 - i)(4 - 6i) + (1 - i)(1)$$

$$= (-1 - 5i) + (2 + 16i) + (-2 - 10i) + (1 - i)$$

= 0 + 0i.

We extend this definition to whole sets by requiring vectors to be pairwise orthogonal. Despite using the same word, careful thought about what objects you are using will eliminate any source of confusion.

Definition OSV Orthogonal Set of Vectors

Suppose that $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is a set of vectors from \mathbb{C}^m . Then S is an **orthogonal** set if every pair of different vectors from S is orthogonal, that is $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ whenever $i \neq j$.

We now define the prototypical orthogonal set, which we will reference repeatedly.

Definition SUV Standard Unit Vectors

Let $\mathbf{e}_j \in \mathbb{C}^m$, $1 \leq j \leq m$ denote the column vectors defined by

$$\left[\mathbf{e}_{j}\right]_{i} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Then the set

$$\{\mathbf{e}_1, \, \mathbf{e}_2, \, \mathbf{e}_3, \, \dots, \, \mathbf{e}_m\} = \{\, \mathbf{e}_j \, | \, 1 \le j \le m\}$$

 \boxtimes

 \wedge

is the set of standard unit vectors in \mathbb{C}^m . (This definition contains Notation SUV.)

Notice that \mathbf{e}_j is identical to column j of the $m \times m$ identity matrix I_m (Definition IM [175]). This observation will often be useful. It is not hard to see that the set of standard unit vectors is an orthogonal set. We will reserve the notation \mathbf{e}_i for these vectors.

Example SUVOS

Standard Unit Vectors are an Orthogonal Set

Compute the inner product of two distinct vectors from the set of standard unit vectors (Definition SUV [409]), say \mathbf{e}_i , \mathbf{e}_j , where $i \neq j$,

$$\langle \mathbf{e}_i, \, \mathbf{e}_j \rangle = \overline{0}0 + \overline{0}0 + \dots + \overline{1}0 + \dots + \overline{0}0 + \dots + \overline{0}1 + \dots + \overline{0}0 + \overline{0}0 = 0(0) + 0(0) + \dots + 1(0) + \dots + 0(1) + \dots + 0(0) + 0(0) = 0$$

So the set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_m\}$ is an orthogonal set.

Example AOS An orthogonal set The set

$$\{\mathbf{x}_{1}, \mathbf{x}_{2}, \mathbf{x}_{3}, \mathbf{x}_{4}\} = \left\{ \begin{bmatrix} 1+i\\1\\1-i\\i \end{bmatrix}, \begin{bmatrix} 1+5i\\6+5i\\-7-i\\1-6i \end{bmatrix}, \begin{bmatrix} -7+34i\\-8-23i\\-10+22i\\30+13i \end{bmatrix}, \begin{bmatrix} -2-4i\\6+i\\4+3i\\6-i \end{bmatrix} \right\}$$

 \boxtimes

 \triangle

is an orthogonal set. Since the inner product is anti-commutative (Theorem IPAC [402]) we can test pairs of different vectors in any order. If the result is zero, then it will also be zero if the inner product is computed in the opposite order. This means there are six different pairs of vectors to use in an inner product computation. We'll do two and you can practice your inner products on the other four.

$$\langle \mathbf{x}_1, \, \mathbf{x}_3 \rangle = (1-i)(-7+34i) + (1)(-8-23i) + (1+i)(-10+22i) + (-i)(30+13i) \\ = (27+41i) + (-8-23i) + (-32+12i) + (13-30i) \\ = 0+0i$$

and

$$\langle \mathbf{x}_2, \, \mathbf{x}_4 \rangle = (1 - 5i)(-2 - 4i) + (6 - 5i)(6 + i) + (-7 + i)(4 + 3i) + (1 + 6i)(6 - i) = (-22 + 6i) + (41 - 24i) + (-31 - 17i) + (12 + 35i) = 0 + 0i$$

 \boxtimes

So far, this section has seen lots of definitions, and lots of theorems establishing unsurprising consequences of those definitions. But here is our first theorem that suggests that inner products and orthogonal vectors have some utility. It is also one of our first illustrations of how to arrive at linear independence as the conclusion of a theorem.

Theorem OSLI

Orthogonal Sets are Linearly Independent

Suppose that S is an orthogonal set of nonzero vectors. Then S is linearly independent. \Box

Proof Let $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ be an orthogonal set of nonzero vectors. To prove the linear independence of S, we can appeal to the definition (Definition LICV [317]) and begin with an arbitrary relation of linear dependence (Definition RLDCV [316]),

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \dots + \alpha_n\mathbf{u}_n = \mathbf{0}.$$

Then, for every $1 \leq i \leq n$, we have

$$\begin{aligned} \alpha_{i} &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} \left(\alpha_{i} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \right\rangle \right) & \text{Theorem PIP [406]} \\ &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} \left(\alpha_{1}(0) + \alpha_{2}(0) + \dots + \alpha_{i} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \right\rangle + \dots + \alpha_{n}(0) \right) & \text{Property ZCN [1597]} \\ &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} \left(\alpha_{1} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{1} \right\rangle + \dots + \alpha_{i} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \right\rangle + \dots + \alpha_{n} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{n} \right\rangle \right) & \text{Definition OSV [409]} \\ &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} \left(\langle \mathbf{u}_{i}, \, \alpha_{1}\mathbf{u}_{1} \rangle + \langle \mathbf{u}_{i}, \, \alpha_{2}\mathbf{u}_{2} \rangle + \dots + \langle \mathbf{u}_{i}, \, \alpha_{n}\mathbf{u}_{n} \rangle \right) & \text{Theorem IPSM [401]} \\ &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} \left\langle \mathbf{u}_{i}, \, \alpha_{1}\mathbf{u}_{1} + \alpha_{2}\mathbf{u}_{2} + \alpha_{3}\mathbf{u}_{3} + \dots + \alpha_{n}\mathbf{u}_{n} \rangle & \text{Theorem IPVA [400]} \\ &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{0} \right\rangle & \text{Definition RLDCV [316]} \\ &= \frac{1}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{i} \rangle} 0 & \text{Definition IP [398]} \end{aligned}$$

= 0

Property ZCN [1597]

So we conclude that $\alpha_i = 0$ for all $1 \le i \le n$ in any relation of linear dependence on S. But this says that S is a linearly independent set since the only way to form a relation of linear dependence is the trivial way (Definition LICV [317]). Boom!

Subsection GSP Gram-Schmidt Procedure

The Gram-Schmidt Procedure is really a theorem. It says that if we begin with a linearly independent set of p vectors, S, then we can do a number of calculations with these vectors and produce an orthogonal set of p vectors, T, so that $\langle S \rangle = \langle T \rangle$. Given the large number of computations involved, it is indeed a procedure to do all the necessary computations, and it is best employed on a computer. However, it also has value in proofs where we may on occasion wish to replace a linearly independent set by an orthogonal set.

This is our first occasion to use the technique of "mathematical induction" for a proof, a technique we will see again several times, especially in Chapter D [882]. So study the simple example described in Technique I [1630] first.

Theorem GSP Gram-Schmidt Procedure

Suppose that $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_p}$ is a linearly independent set of vectors in \mathbb{C}^m . Define

the vectors \mathbf{u}_i , $1 \leq i \leq p$ by

$$\mathbf{u}_{i} = \mathbf{v}_{i} - \frac{\langle \mathbf{u}_{1}, \, \mathbf{v}_{i} \rangle}{\langle \mathbf{u}_{1}, \, \mathbf{u}_{1} \rangle} \mathbf{u}_{1} - \frac{\langle \mathbf{u}_{2}, \, \mathbf{v}_{i} \rangle}{\langle \mathbf{u}_{2}, \, \mathbf{u}_{2} \rangle} \mathbf{u}_{2} - \frac{\langle \mathbf{u}_{3}, \, \mathbf{v}_{i} \rangle}{\langle \mathbf{u}_{3}, \, \mathbf{u}_{3} \rangle} \mathbf{u}_{3} - \dots - \frac{\langle \mathbf{u}_{i-1}, \, \mathbf{v}_{i} \rangle}{\langle \mathbf{u}_{i-1}, \, \mathbf{u}_{i-1} \rangle} \mathbf{u}_{i-1}$$

Then if $T = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_p}$, then T is an orthogonal set of non-zero vectors, and $\langle T \rangle = \langle S \rangle$.

Proof We will prove the result by using induction on p (Technique I [1630]). To begin, we prove that T has the desired properties when p = 1. In this case $\mathbf{u}_1 = \mathbf{v}_1$ and $T = {\mathbf{u}_1} = {\mathbf{v}_1} = S$. Because S and T are equal, $\langle S \rangle = \langle T \rangle$. Equally trivial, T is an orthogonal set. If $\mathbf{u}_1 = \mathbf{0}$, then S would be a linearly dependent set, a contradiction.

Suppose that the theorem is true for any set of p-1 linearly independent vectors. Let $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_p}$ be a linearly independent set of p vectors. Then $S' = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{p-1}}$ is also linearly independent. So we can apply the theorem to S' and construct the vectors $T' = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_{p-1}}$. T' is therefore an orthogonal set of nonzero vectors and $\langle S' \rangle = \langle T' \rangle$. Define

$$\mathbf{u}_p = \mathbf{v}_p - \frac{\langle \mathbf{u}_1, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_1, \, \mathbf{u}_1 \rangle} \mathbf{u}_1 - \frac{\langle \mathbf{u}_2, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_2, \, \mathbf{u}_2 \rangle} \mathbf{u}_2 - \frac{\langle \mathbf{u}_3, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_3, \, \mathbf{u}_3 \rangle} \mathbf{u}_3 - \dots - \frac{\langle \mathbf{u}_{p-1}, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_{p-1}, \, \mathbf{u}_{p-1} \rangle} \mathbf{u}_{p-1}$$

and let $T = T' \cup {\mathbf{u}_p}$. We need to now show that T has several properties by building on what we know about T'. But first notice that the above equation has no problems with the denominators $(\langle \mathbf{u}_i, \mathbf{u}_i \rangle)$ being zero, since the \mathbf{u}_i are from T', which is composed of nonzero vectors.

We show that $\langle T \rangle = \langle S \rangle$, by first establishing that $\langle T \rangle \subseteq \langle S \rangle$. Suppose $\mathbf{x} \in \langle T \rangle$, so

$$\mathbf{x} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_p \mathbf{u}_p$$

The term $a_p \mathbf{u}_p$ is a linear combination of vectors from T' and the vector \mathbf{v}_p , while the remaining terms are a linear combination of vectors from T'. Since $\langle T' \rangle = \langle S' \rangle$, any term that is a multiple of a vector from T' can be rewritten as a linear combination of vectors from S'. The remaining term $a_p \mathbf{v}_p$ is a multiple of a vector in S. So we see that \mathbf{x} can be rewritten as a linear combination of vectors from S, i.e. $\mathbf{x} \in \langle S \rangle$.

To show that $\langle S \rangle \subseteq \langle T \rangle$, begin with $\mathbf{y} \in \langle S \rangle$, so

$$\mathbf{y} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_p \mathbf{v}_p$$

Rearrange our defining equation for \mathbf{u}_p by solving for \mathbf{v}_p . Then the term $a_p\mathbf{v}_p$ is a multiple of a linear combination of elements of T. The remaining terms are a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{p-1}$, hence an element of $\langle S' \rangle = \langle T' \rangle$. Thus these remaining terms can be written as a linear combination of the vectors in T'. So \mathbf{y} is a linear combination of vectors from T, i.e. $\mathbf{y} \in \langle T \rangle$.

The elements of T' are nonzero, but what about \mathbf{u}_p ? Suppose to the contrary that $\mathbf{u}_p = \mathbf{0}$,

$$0 = \mathbf{u}_p = \mathbf{v}_p - \frac{\langle \mathbf{u}_1, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_1, \, \mathbf{u}_1 \rangle} \mathbf{u}_1 - \frac{\langle \mathbf{u}_2, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_2, \, \mathbf{u}_2 \rangle} \mathbf{u}_2 - \frac{\langle \mathbf{u}_3, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_3, \, \mathbf{u}_3 \rangle} \mathbf{u}_3 - \dots - \frac{\langle \mathbf{u}_{p-1}, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_{p-1}, \, \mathbf{u}_{p-1} \rangle} \mathbf{u}_{p-1}$$
$$\mathbf{v}_p = \frac{\langle \mathbf{u}_1, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_1, \, \mathbf{u}_1 \rangle} \mathbf{u}_1 + \frac{\langle \mathbf{u}_2, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_2, \, \mathbf{u}_2 \rangle} \mathbf{u}_2 + \frac{\langle \mathbf{u}_3, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_3, \, \mathbf{u}_3 \rangle} \mathbf{u}_3 + \dots + \frac{\langle \mathbf{u}_{p-1}, \, \mathbf{v}_p \rangle}{\langle \mathbf{u}_{p-1}, \, \mathbf{u}_{p-1} \rangle} \mathbf{u}_{p-1}$$

Version 2.99

Since $\langle S' \rangle = \langle T' \rangle$ we can write the vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_{p-1}$ on the right side of this equation in terms of the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{p-1}$ and we then have the vector \mathbf{v}_p expressed as a linear combination of the other p-1 vectors in S, implying that S is a linearly dependent set (Theorem DLDS [365]), contrary to our lone hypothesis about S.

Finally, it is a simple matter to establish that T is an orthogonal set, though it will not appear so simple looking. Think about your objects as you work through the following what is a vector and what is a scalar. Since T' is an orthogonal set by induction, most pairs of elements in T are already known to be orthogonal. We just need to test "new" inner products, between \mathbf{u}_p and \mathbf{u}_i , for $1 \leq i \leq p - 1$. Here we go, using summation notation,

$$\begin{split} \langle \mathbf{u}_{i}, \, \mathbf{u}_{p} \rangle &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} - \sum_{k=1}^{p-1} \frac{\langle \mathbf{u}_{k}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \rangle} \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \left\langle \mathbf{u}_{i}, \sum_{k=1}^{p-1} \frac{\langle \mathbf{u}_{k}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \rangle} \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \sum_{k=1}^{p-1} \left\langle \mathbf{u}_{i}, \, \frac{\langle \mathbf{u}_{k}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \rangle} \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \sum_{k=1}^{p-1} \frac{\langle \mathbf{u}_{k}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \rangle} \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \sum_{k=1}^{p-1} \frac{\langle \mathbf{u}_{k}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \right\rangle - \frac{\langle \mathbf{u}_{i}, \, \mathbf{v}_{p} \rangle}{\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \rangle} \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{i}, \, \mathbf{u}_{k} \right\rangle \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k} \right\rangle \\ &= \left\langle \mathbf{u}_{k}, \, \mathbf{u}_{k}$$

$$= \langle \mathbf{u}_i, \, \mathbf{v}_p \rangle - \langle \mathbf{u}_i, \, \mathbf{v}_p \rangle - \sum_{k \neq i} 0$$
$$= 0$$

Example GSTV Gram-Schmidt of three vectors

We will illustrate the Gram-Schmidt process with three vectors. Begin with the linearly independent (check this!) set

$$S = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3\} = \left\{ \begin{bmatrix} 1\\1+i\\1 \end{bmatrix}, \, \begin{bmatrix} -i\\1\\1+i \end{bmatrix}, \, \begin{bmatrix} 0\\i\\i \end{bmatrix} \right\}$$

Then

$$\mathbf{u}_{1} = \mathbf{v}_{1} = \begin{bmatrix} 1\\1+i\\1 \end{bmatrix}$$
$$\mathbf{u}_{2} = \mathbf{v}_{2} - \frac{\langle \mathbf{u}_{1}, \mathbf{v}_{2} \rangle}{\langle \mathbf{u}_{1}, \mathbf{u}_{1} \rangle} \mathbf{u}_{1} = \frac{1}{4} \begin{bmatrix} -2-3i\\1-i\\2+5i \end{bmatrix}$$

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{u}_1, \, \mathbf{v}_3 \rangle}{\langle \mathbf{u}_1, \, \mathbf{u}_1 \rangle} \mathbf{u}_1 - \frac{\langle \mathbf{u}_2, \, \mathbf{v}_3 \rangle}{\langle \mathbf{u}_2, \, \mathbf{u}_2 \rangle} \mathbf{u}_2 = \frac{1}{11} \begin{bmatrix} -3 - i \\ 1 + 3i \\ -1 - i \end{bmatrix}$$

and

$$T = \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 1\\1+i\\1 \end{bmatrix}, \, \frac{1}{4} \begin{bmatrix} -2-3i\\1-i\\2+5i \end{bmatrix}, \, \frac{1}{11} \begin{bmatrix} -3-i\\1+3i\\-1-i \end{bmatrix} \right\}$$

is an orthogonal set (which you can check) of nonzero vectors and $\langle T \rangle = \langle S \rangle$ (all by Theorem GSP [413]). Of course, as a by-product of orthogonality, the set T is also linearly independent (Theorem OSLI [412]). \boxtimes

One final definition related to orthogonal vectors.

Definition ONS OrthoNormal Set

Suppose $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is an orthogonal set of vectors such that $||\mathbf{u}_i|| = 1$ for all $1 \leq i \leq n$. Then S is an **orthonormal** set of vectors.

Once you have an orthogonal set, it is easy to convert it to an orthonormal set — multiply each vector by the reciprocal of its norm, and the resulting vector will have norm 1. This scaling of each vector will not affect the orthogonality properties (apply Theorem IPSM [401]).

Example ONTV Orthonormal set, three vectors \land

The set

$$T = \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 1\\1+i\\1 \end{bmatrix}, \, \frac{1}{4} \begin{bmatrix} -2-3i\\1-i\\2+5i \end{bmatrix}, \, \frac{1}{11} \begin{bmatrix} -3-i\\1+3i\\-1-i \end{bmatrix} \right\}$$

from Example GSTV [417] is an orthogonal set. We compute the norm of each vector,

$$\|\mathbf{u}_1\| = 2$$
 $\|\mathbf{u}_2\| = \frac{1}{2}\sqrt{11}$ $\|\mathbf{u}_3\| = \frac{\sqrt{2}}{\sqrt{11}}$

Converting each vector to a norm of 1, yields an orthonormal set,

$$\mathbf{w}_{1} = \frac{1}{2} \begin{bmatrix} 1\\1+i\\1 \end{bmatrix}$$
$$\mathbf{w}_{2} = \frac{1}{\frac{1}{2}\sqrt{11}} \frac{1}{4} \begin{bmatrix} -2-3i\\1-i\\2+5i \end{bmatrix} = \frac{1}{2\sqrt{11}} \begin{bmatrix} -2-3i\\1-i\\2+5i \end{bmatrix}$$
$$\mathbf{w}_{3} = \frac{1}{\frac{\sqrt{2}}{\sqrt{11}}} \frac{1}{11} \begin{bmatrix} -3-i\\1+3i\\-1-i \end{bmatrix} = \frac{1}{\sqrt{22}} \begin{bmatrix} -3-i\\1+3i\\-1-i \end{bmatrix}$$

Example ONFV Orthonormal set, four vectors As an exercise convert the linearly independent set

$$S = \left\{ \begin{bmatrix} 1+i\\1\\1-i\\i \end{bmatrix}, \begin{bmatrix} i\\1+i\\-1\\-1\\-i \end{bmatrix}, \begin{bmatrix} i\\-i\\-1+i\\1 \end{bmatrix}, \begin{bmatrix} -1-i\\i\\1\\-1 \end{bmatrix} \right\}$$

to an orthogonal set via the Gram-Schmidt Process (Theorem GSP [413]) and then scale the vectors to norm 1 to create an orthonormal set. You should get the same set you would if you scaled the orthogonal set of Example AOS [410] to become an orthonormal set. \boxtimes

It is crazy to do all but the simplest and smallest instances of the Gram-Schmidt procedure by hand. Well, OK, maybe just once or twice to get a good understanding of Theorem GSP [413]. After that, let a machine do the work for you. That's what they are for. See: Computation GSP.MMA [1573]

We will see orthonormal sets again in Subsection MINM.UM [544]. They are intimately related to unitary matrices (Definition UM [544]) through Theorem CUMOS [547]. Some of the utility of orthonormal sets is captured by Theorem COB [793] in Subsection B.OBC [792]. Orthonormal sets appear once again in Section OD [1414] where they are key in orthonormal diagonalization.

Subsection READ Reading Questions

1. Is the set

$$\left\{ \begin{bmatrix} 1\\-1\\2 \end{bmatrix}, \begin{bmatrix} 5\\3\\-1 \end{bmatrix}, \begin{bmatrix} 8\\4\\-2 \end{bmatrix} \right\}$$

an orthogonal set? Why?

- 2. What is the distinction between an orthogonal set and an orthonormal set?
- 3. What is nice about the output of the Gram-Schmidt process?

Subsection EXC Exercises

C20 Complete Example AOS [410] by verifying that the four remaining inner products are zero.

Contributed by Robert Beezer

C21 Verify that the set T created in Example GSTV [417] by the Gram-Schmidt Procedure is an orthogonal set. Contributed by Robert Beezer

M60 Suppose that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\} \subseteq \mathbb{C}^n$ is an orthonormal set. Prove that $\mathbf{u} + \mathbf{v}$ is not orthogonal to $\mathbf{v} + \mathbf{w}$. Contributed by Manley Perkel

T10 Prove part 2 of the conclusion of Theorem IPVA [400]. Contributed by Robert Beezer

T11 Prove part 2 of the conclusion of Theorem IPSM [401]. Contributed by Robert Beezer

T20 Suppose that $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{C}^n, \alpha, \beta \in \mathbb{C}$ and \mathbf{u} is orthogonal to both \mathbf{v} and \mathbf{w} . Prove that \mathbf{u} is orthogonal to $\alpha \mathbf{v} + \beta \mathbf{w}$.

Contributed by Robert Beezer Solution [424]

T30 Suppose that the set S in the hypothesis of Theorem GSP [413] is not just linearly independent, but is also orthogonal. Prove that the set T created by the Gram-Schmidt procedure is equal to S. (Note that we are getting a stronger conclusion than $\langle T \rangle = \langle S \rangle$ — the conclusion is that T = S.) In other words, it is pointless to apply the Gram-Schmidt procedure to a set that is already orthogonal.

Contributed by Steve Canfield

Subsection SOL Solutions

T20 Contributed by Robert Beezer Statement [422] Vectors are orthogonal if their inner product is zero (Definition OV [408]), so we compute,

$$\langle \mathbf{u}, \, \alpha \mathbf{v} + \beta \mathbf{w} \rangle = \langle \mathbf{u}, \, \alpha \mathbf{v} \rangle + \langle \mathbf{u}, \, \beta \mathbf{w} \rangle$$

$$= \alpha \langle \mathbf{u}, \, \mathbf{v} \rangle + \beta \langle \mathbf{u}, \, \mathbf{w} \rangle$$

$$= \alpha \langle \mathbf{0} \rangle + \beta \langle \mathbf{0} \rangle$$

$$= 0$$
Theorem IPVA [400]
Theorem IPSM [401]
Definition OV [408]
Theorem IPVA [400]
Theorem IPVA [401]
Theorem IPVA [401

So by Definition OV [408], **u** and $\alpha \mathbf{v} + \beta \mathbf{w}$ are an orthogonal pair of vectors.

Chapter M Matrices

We have made frequent use of matrices for solving systems of equations, and we have begun to investigate a few of their properties, such as the null space and nonsingularity. In this chapter, we will take a more systematic approach to the study of matrices.

Section MO Matrix Operations

In this section we will back up and start simple. First a definition of a totally general set of matrices.

Definition VSM Vector Space of $m \times n$ Matrices

The vector space M_{mn} is the set of all $m \times n$ matrices with entries from the set of complex numbers.

(This definition contains Notation VSM.)

Subsection MEASM Matrix Equality, Addition, Scalar Multiplication

Just as we made, and used, a careful definition of equality for column vectors, so too, we have precise definitions for matrices.

Definition ME

 \triangle

Matrix Equality

The $m \times n$ matrices A and B are **equal**, written A = B provided $[A]_{ij} = [B]_{ij}$ for all $1 \le i \le m, 1 \le j \le n$. (This definition contains Notation ME.)

So equality of matrices translates to the equality of complex numbers, on an entry-byentry basis. Notice that we now have yet another definition that uses the symbol "=" for shorthand. Whenever a theorem has a conclusion saying two matrices are equal (think about your objects), we will consider appealing to this definition as a way of formulating the toplevel structure of the proof. We will now define two operations on the set M_{mn} . Again, we will overload a symbol ('+') and a convention (juxtaposition for scalar multiplication).

Definition MA Matrix Addition

Given the $m \times n$ matrices A and B, define the **sum** of A and B as an $m \times n$ matrix, written A + B, according to

$$[A+B]_{ij} = [A]_{ij} + [B]_{ij} \qquad 1 \le i \le m, \ 1 \le j \le n$$

(This definition contains Notation MA.)

So matrix addition takes two matrices of the same size and combines them (in a natural way!) to create a new matrix of the same size. Perhaps this is the "obvious" thing to do, but it doesn't relieve us from the obligation to state it carefully.

 \triangle

Example MA Addition of two matrices in M_{23} If

$$A = \begin{bmatrix} 2 & -3 & 4 \\ 1 & 0 & -7 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 6 & 2 & -4 \\ 3 & 5 & 2 \end{bmatrix}$$

then

$$A + B = \begin{bmatrix} 2 & -3 & 4 \\ 1 & 0 & -7 \end{bmatrix} + \begin{bmatrix} 6 & 2 & -4 \\ 3 & 5 & 2 \end{bmatrix} = \begin{bmatrix} 2+6 & -3+2 & 4+(-4) \\ 1+3 & 0+5 & -7+2 \end{bmatrix} = \begin{bmatrix} 8 & -1 & 0 \\ 4 & 5 & -5 \end{bmatrix}$$

Our second operation takes two objects of different types, specifically a number and a matrix, and combines them to create another matrix. As with vectors, in this context we call a number a **scalar** in order to emphasize that it is not a matrix.

Definition MSM Matrix Scalar Multiplication

Given the $m \times n$ matrix A and the scalar $\alpha \in \mathbb{C}$, the scalar multiple of A is an $m \times n$ matrix, written αA and defined according to

$$[\alpha A]_{ij} = \alpha [A]_{ij} \qquad 1 \le i \le m, \ 1 \le j \le n$$

 \boxtimes

(This definition contains Notation MSM.)

Notice again that we have yet another kind of multiplication, and it is again written putting two symbols side-by-side. Computationally, scalar matrix multiplication is very easy.

Example MSM Scalar multiplication in M_{32} If

$$A = \begin{bmatrix} 2 & 8\\ -3 & 5\\ 0 & 1 \end{bmatrix}$$

and $\alpha = 7$, then

$$\alpha A = 7 \begin{bmatrix} 2 & 8 \\ -3 & 5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7(2) & 7(8) \\ 7(-3) & 7(5) \\ 7(0) & 7(1) \end{bmatrix} = \begin{bmatrix} 14 & 56 \\ -21 & 35 \\ 0 & 7 \end{bmatrix}$$

 \boxtimes

 \triangle

Subsection VSP Vector Space Properties

With definitions of matrix addition and scalar multiplication we can now state, and prove, several properties of each operation, and some properties that involve their interplay. We now collect ten of them here for later reference.

Theorem VSPM

Vector Space Properties of Matrices

Suppose that M_{mn} is the set of all $m \times n$ matrices (Definition VSM [426]) with addition and scalar multiplication as defined in Definition MA [427] and Definition MSM [428]. Then

- ACM Additive Closure, Matrices If $A, B \in M_{mn}$, then $A + B \in M_{mn}$.
- SCM Scalar Closure, Matrices If $\alpha \in \mathbb{C}$ and $A \in M_{mn}$, then $\alpha A \in M_{mn}$.
- CM Commutativity, Matrices If $A, B \in M_{mn}$, then A + B = B + A.
- AAM Additive Associativity, Matrices If $A, B, C \in M_{mn}$, then A + (B + C) = (A + B) + C.

• ZM Zero Vector, Matrices

There is a matrix, \mathcal{O} , called the **zero matrix**, such that $A + \mathcal{O} = A$ for all $A \in M_{mn}$.

- AIM Additive Inverses, Matrices If $A \in M_{mn}$, then there exists a matrix $-A \in M_{mn}$ so that $A + (-A) = \mathcal{O}$.
- SMAM Scalar Multiplication Associativity, Matrices If $\alpha, \beta \in \mathbb{C}$ and $A \in M_{mn}$, then $\alpha(\beta A) = (\alpha \beta)A$.
- DMAM Distributivity across Matrix Addition, Matrices If $\alpha \in \mathbb{C}$ and $A, B \in M_{mn}$, then $\alpha(A + B) = \alpha A + \alpha B$.
- DSAM Distributivity across Scalar Addition, Matrices If $\alpha, \beta \in \mathbb{C}$ and $A \in M_{mn}$, then $(\alpha + \beta)A = \alpha A + \beta A$.
- OM One, Matrices If $A \in M_{mn}$, then 1A = A.

 \square

Proof While some of these properties seem very obvious, they all require proof. However, the proofs are not very interesting, and border on tedious. We'll prove one version of distributivity very carefully, and you can test your proof-building skills on some of the others. We'll give our new notation for matrix entries a workout here. Compare the style of the proofs here with those given for vectors in Theorem VSPCV [205] — while the objects here are more complicated, our notation makes the proofs cleaner.

To prove Property DSAM [431], $(\alpha + \beta)A = \alpha A + \beta A$, we need to establish the equality of two matrices (see Technique GS [1617]). Definition ME [427] says we need to establish the equality of their entries, one-by-one. How do we do this, when we do not even know how many entries the two matrices might have? This is where Notation ME [427] comes into play. Ready? Here we go.

For any i and j, $1 \le i \le m$, $1 \le j \le n$,

$\left[(\alpha + \beta)A \right]_{ij} = (\alpha + \beta) \left[A\right]_{ij}$	Definition MSM [428]
$= \alpha \left[A \right]_{ij} + \beta \left[A \right]_{ij}$	Distributivity in $\mathbb C$
$= \left[\alpha A\right]_{ij} + \left[\beta A\right]_{ij}$	Definition MSM [428]
$= \left[\alpha A + \beta A\right]_{ij}$	Definition MA [427]

There are several things to notice here. (1) Each equals sign is an equality of numbers. (2) The two ends of the equation, being true for any i and j, allow us to conclude the equality of the matrices by Definition ME [427]. (3) There are several plus signs, and several instances of juxtaposition. Identify each one, and state exactly what operation is being represented by each.

For now, note the similarities between Theorem VSPM [430] about matrices and Theorem VSPCV [205] about vectors.

The zero matrix described in this theorem, \mathcal{O} , is what you would expect — a matrix full of zeros.

Definition ZM

Zero Matrix

The $m \times n$ zero matrix is written as $\mathcal{O} = \mathcal{O}_{m \times n}$ and defined by $[\mathcal{O}]_{ij} = 0$, for all $1 \le i \le m$, $1 \le j \le n$. (This definition contains Notation ZM.)

Subsection TSM Transposes and Symmetric Matrices

We describe one more common operation we can perform on matrices. Informally, to transpose a matrix is to build a new matrix by swapping its rows and columns.

Definition TM Transpose of a Matrix

Given an $m \times n$ matrix A, its **transpose** is the $n \times m$ matrix A^t given by

$$\left[A^t\right]_{ij} = [A]_{ji}\,,\quad 1\leq i\leq n,\, 1\leq j\leq m.$$

(This definition contains Notation TM.)

Example TM Transpose of a 3×4 matrix \wedge

Suppose

$$D = \begin{bmatrix} 3 & 7 & 2 & -3 \\ -1 & 4 & 2 & 8 \\ 0 & 3 & -2 & 5 \end{bmatrix}.$$

We could formulate the transpose, entry-by-entry, using the definition. But it is easier to just systematically rewrite rows as columns (or vice-versa). The form of the definition given will be more useful in proofs. So we have

$$D^{t} = \begin{bmatrix} 3 & -1 & 0 \\ 7 & 4 & 3 \\ 2 & 2 & -2 \\ -3 & 8 & 5 \end{bmatrix}$$

 \boxtimes

It will sometimes happen that a matrix is equal to its transpose. In this case, we will call a matrix **symmetric**. These matrices occur naturally in certain situations, and also have some nice properties, so it is worth stating the definition carefully. Informally a matrix is symmetric if we can "flip" it about the main diagonal (upper-left corner, running down to the lower-right corner) and have it look unchanged.

Definition SYM Symmetric Matrix The matrix A is symmetric if $A = A^t$.

Example SYM

 \wedge

A symmetric 5×5 matrix The matrix

$$E = \begin{bmatrix} 2 & 3 & -9 & 5 & 7 \\ 3 & 1 & 6 & -2 & -3 \\ -9 & 6 & 0 & -1 & 9 \\ 5 & -2 & -1 & 4 & -8 \\ 7 & -3 & 9 & -8 & -3 \end{bmatrix}$$

is symmetric.

You might have noticed that Definition SYM [434] did not specify the size of the matrix A, as has been our custom. That's because it wasn't necessary. An alternative would have been to state the definition just for square matrices, but this is the substance of the next proof. Before reading the next proof, we want to offer you some advice about how to become more proficient at constructing proofs. Perhaps you can apply this advice to the next theorem. Have a peek at Technique P [1633] now.

Theorem SMS Symmetric Matrices are Square

Suppose that A is a symmetric matrix. Then A is square.

Proof We start by specifying A's size, without assuming it is square, since we are trying to *prove* that, so we can't also assume it. Suppose A is an $m \times n$ matrix. Because A is symmetric, we know by Definition SM [893] that $A = A^t$. So, in particular, Definition ME [427] requires that A and A^t must have the same size. The size of A^t is $n \times m$. Because A

 \boxtimes

 \square

has m rows and A^t has n rows, we conclude that m = n, and hence A must be square by Definition SQM [173].

We finish this section with three easy theorems, but they illustrate the interplay of our three new operations, our new notation, and the techniques used to prove matrix equalities.

Theorem TMA Transpose and Matrix Addition

Suppose that A and B are $m \times n$ matrices. Then $(A + B)^t = A^t + B^t$.

Proof The statement to be proved is an equality of matrices, so we work entry-by-entry and use Definition ME [427]. Think carefully about the objects involved here, and the many uses of the plus sign. For $1 \le i \le m$, $1 \le j \le n$,

$$\begin{bmatrix} (A+B)^t \end{bmatrix}_{ij} = \begin{bmatrix} A+B \end{bmatrix}_{ji} & \text{Definition TM [433]} \\ = \begin{bmatrix} A \end{bmatrix}_{ji} + \begin{bmatrix} B \end{bmatrix}_{ji} & \text{Definition MA [427]} \\ = \begin{bmatrix} A^t \end{bmatrix}_{ij} + \begin{bmatrix} B^t \end{bmatrix}_{ij} & \text{Definition TM [433]} \\ = \begin{bmatrix} A^t + B^t \end{bmatrix}_{ij} & \text{Definition MA [427]} \end{bmatrix}$$

Since the matrices $(A + B)^t$ and $A^t + B^t$ agree at each entry, Definition ME [427] tells us the two matrices are equal.

Theorem TMSM Transpose and Matrix Scalar Multiplication Suppose that $\alpha \in \mathbb{C}$ and A is an $m \times n$ matrix. Then $(\alpha A)^t = \alpha A^t$.

Proof The statement to be proved is an equality of matrices, so we work entry-by-entry and use Definition ME [427]. Notice that the desired equality is of $n \times m$ matrices, and think carefully about the objects involved here, plus the many uses of juxtaposition. For $1 \le i \le m, 1 \le j \le n$,

$\left[(\alpha A)^t \right]_{ji} = [\alpha A]_{ij}$	Definition TM [433]
$= \alpha \left[A \right]_{ij}$	Definition MSM [428]
$= \alpha \left[A^t \right]_{ji}$	Definition TM [433]
$= \left[\alpha A^t \right]_{ji}$	Definition MSM [428]

Since the matrices $(\alpha A)^t$ and αA^t agree at each entry, Definition ME [427] tells us the two matrices are equal.

Theorem TT Transpose of a Transpose

Suppose that A is an $m \times n$ matrix. Then $(A^t)^t = A$.

Proof We again want to prove an equality of matrices, so we work entry-by-entry and use Definition ME [427]. For $1 \le i \le m$, $1 \le j \le n$,

$$\left[\left(A^t \right)^t \right]_{ij} = \left[A^t \right]_{ji} \qquad \text{Definition TM [433]}$$

 \square

 $= [A]_{ij}$

Definition TM [433]

It's usually straightforward to coax the transpose of a matrix out of a computational device. See: Computation TM.MMA [1575] Computation TM.TI86 [1580] Computation TM.SAGE [1591]

Subsection MCC Matrices and Complex Conjugation

As we did with vectors (Definition CCCV [396]), we can define what it means to take the conjugate of a matrix.

Definition CCM Complex Conjugate of a Matrix

Suppose A is an $m \times n$ matrix. Then the **conjugate** of A, written \overline{A} is an $m \times n$ matrix defined by

$$\left[\overline{A}\right]_{ij} = \overline{[A]_{ij}}$$

(This definition contains Notation CCM.)

Example CCM

 \wedge

Complex conjugate of a matrix If

$$A = \begin{bmatrix} 2 - i & 3 & 5 + 4i \\ -3 + 6i & 2 - 3i & 0 \end{bmatrix}$$

then

$$\overline{A} = \begin{bmatrix} 2+i & 3 & 5-4i \\ -3-6i & 2+3i & 0 \end{bmatrix}$$

 \square

The interplay between the conjugate of a matrix and the two operations on matrices is what you might expect.

Theorem CRMA Conjugation Respects Matrix Addition

Suppose that A and B are $m \times n$ matrices. Then $\overline{A + B} = \overline{A} + \overline{B}$.

Proof For $1 \le i \le m, 1 \le j \le n$,

$$\overline{A+B}]_{ij} = \overline{[A+B]_{ij}}$$
$$= \overline{[A]_{ij} + [B]_{ij}}$$
$$= \overline{[A]_{ij} + \overline{[B]_{ij}}}$$
$$= \overline{[A]_{ij} + \overline{[B]_{ij}}}$$
$$= \overline{[A]_{ij} + \overline{[B]_{ij}}}$$
$$= \overline{[A+\overline{B}]_{ij}}$$

Definition CCM [438] Definition MA [427] Theorem CCRA [1599] Definition CCM [438] Definition MA [427] Since the matrices $\overline{A + B}$ and $\overline{A} + \overline{B}$ are equal in each entry, Definition ME [427] says that $\overline{A + B} = \overline{A} + \overline{B}$.

Theorem CRMSM

Conjugation Respects Matrix Scalar Multiplication \overline{C}

Suppose that $\alpha \in \mathbb{C}$ and A is an $m \times n$ matrix. Then $\overline{\alpha A} = \overline{\alpha}\overline{A}$.

Proof For $1 \le i \le m$, $1 \le j \le n$, $\left[\overline{\alpha A}\right]_{ij} = \overline{\left[\alpha A\right]_{ij}}$ Defin $= \overline{\alpha \left[A\right]_{ij}}$ Defin $= \overline{\alpha} \overline{\left[A\right]_{ij}}$ Theo $= \overline{\alpha} \overline{\left[A\right]_{ij}}$ Defin $= \left[\overline{\alpha} \overline{A}\right]_{ij}$ Defin

Definition CCM [438] Definition MSM [428] Theorem CCRM [1599] Definition CCM [438] Definition MSM [428]

Since the matrices $\overline{\alpha A}$ and $\overline{\alpha}\overline{A}$ are equal in each entry, Definition ME [427] says that $\overline{\alpha A} = \overline{\alpha}\overline{A}$.

Theorem CCM Conjugate of the Conjugate of a Matrix Suppose that A is an $m \times n$ matrix. Then $\overline{(\overline{A})} = A$. Proof For $1 \le i \le m, 1 \le j \le n,$ $\overline{\left[\overline{(\overline{A})}\right]_{ij}} = \overline{[\overline{A}]_{ij}}$ Definition CCM [438]

$$= \overline{[A]_{ij}}$$

$$= [A]_{ij}$$
Definition CCM [438]
Theorem CCT [1600]

Since the matrices $\overline{(A)}$ and A are equal in each entry, Definition ME [427] says that $\overline{(A)} = A$.

Finally, we will need the following result about matrix conjugation and transposes later. **Theorem MCT Matrix Conjugation and Transposes**

Suppose that A is an $m \times n$ matrix. Then $\overline{(A^t)} = (\overline{A})^t$.

Proof For $1 \le i \le m, 1 \le j \le n$,

$$\overline{(A^{t})}\Big]_{ji} = \overline{[A^{t}]_{ji}} \qquad \text{Definition CCM [438]} \\
= \overline{[A]_{ij}} \qquad \text{Definition TM [433]} \\
= \left[\overline{A}\right]_{ij} \qquad \text{Definition CCM [438]} \\
= \left[\left(\overline{A}\right)^{t}\right]_{ii} \qquad \text{Definition TM [433]}$$

Since the matrices $\overline{(A^t)}$ and $(\overline{A})^t$ are equal in each entry, Definition ME [427] says that $\overline{(A^t)} = (\overline{A})^t$.

Subsection AM Adjoint of a Matrix

The combination of transposing and conjugating a matrix will be important in subsequent sections, such as Subsection MINM.UM [544] and Section OD [1414]. We make a key definition here and prove some basic results in the same spirit as those above.

Definition A

Adjoint

If A is a matrix, then its **adjoint** is $A^* = (\overline{A})^t$. (This definition contains Notation A.)

You will see the adjoint written elsewhere variously as A^H , A^* or A^{\dagger} . Notice that Theorem MCT [441] says it does not really matter if we conjugate and then transpose, or transpose and then conjugate.

Theorem AMA Adjoint and Matrix Addition

Suppose A and B are matrices of the same size. Then $(A + B)^* = A^* + B^*$.

Proof

$$(A+B)^* = (\overline{A+B})^t \qquad \text{Definition A [442]} \\ = (\overline{A}+\overline{B})^t \qquad \text{Theorem CRMA [439]}$$

 \triangle

[436]

$$= (\overline{A})^{t} + (\overline{B})^{t}$$
Theorem TMA [436]

$$= A^{*} + B^{*}$$
Definition A [442]

Theorem AMSM Adjoint and Matrix Scalar Multiplication

Suppose $\alpha \in \mathbb{C}$ is a scalar and A is a matrix. Then $(\alpha A)^* = \overline{\alpha} A^*$.

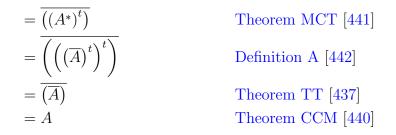
Proof

$$(\alpha A)^* = (\overline{\alpha A})^t \qquad \text{Definition A [442]} \\ = (\overline{\alpha}\overline{A})^t \qquad \text{Theorem CRMSM [440]} \\ = \overline{\alpha}(\overline{A})^t \qquad \text{Theorem TMSM [436]} \\ = \overline{\alpha}A^* \qquad \text{Definition A [442]}$$

Theorem AA Adjoint of an Adjoint Suppose that A is a matrix. Then $(A^*)^* = A$ Proof

$$(A^*)^* = \left(\overline{(A^*)}\right)^t$$
 Definition A [442]

 \square



Take note of how the theorems in this section, while simple, build on earlier theorems and definitions and never descend to the level of entry-by-entry proofs based on Definition ME [427]. In other words, the equal signs that appear in the previous proofs are equalities of matrices, not scalars (which is the opposite of a proof like that of Theorem TMA [436]).

Subsection READ Reading Questions

1. Perform the following matrix computation.

$$(6)\begin{bmatrix}2 & -2 & 8 & 1\\4 & 5 & -1 & 3\\7 & -3 & 0 & 2\end{bmatrix} + (-2)\begin{bmatrix}2 & 7 & 1 & 2\\3 & -1 & 0 & 5\\1 & 7 & 3 & 3\end{bmatrix}$$

- 2. Theorem VSPM [430] reminds you of what previous theorem? How strong is the similarity?
- 3. Compute the transpose of the matrix below.

$$\begin{bmatrix} 6 & 8 & 4 \\ -2 & 1 & 0 \\ 9 & -5 & 6 \end{bmatrix}$$

Subsection EXC Exercises

C10 Let
$$A = \begin{bmatrix} 1 & 4 & -3 \\ 6 & 3 & 0 \end{bmatrix}$$
, $B = \begin{bmatrix} 3 & 2 & 1 \\ -2 & -6 & 5 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 4 \\ 4 & 0 \\ -2 & 2 \end{bmatrix}$. Let $\alpha = 4$ and $\beta = 1/2$.

Perform the following calculations:

- 1. A + B
- 2. A + C
- 3. $B^t + C$
- 4. $A + B^t$
- 5. βC
- 6. 4A 3B
- 7. $A^t + \alpha C$
- 8. $A + B C^t$
- 9. $4A + 2B 5C^t$

Contributed by Chris Black Solution [453]

C11 Solve the given vector equation for x, or explain why no solution exists:

$$2\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 2 \end{bmatrix} - 3\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & x \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 5 & -2 \end{bmatrix}$$

Contributed by Chris Black Solution [454]

C12 Solve the given matrix equation for α , or explain why no solution exists:

$$\alpha \begin{bmatrix} 1 & 3 & 4 \\ 2 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 4 & 3 & -6 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 7 & 12 & 6 \\ 6 & 4 & -2 \end{bmatrix}$$

Contributed by Chris Black Solution [454]

C13 Solve the given matrix equation for α , or explain why no solution exists:

$$\alpha \begin{bmatrix} 3 & 1 \\ 2 & 0 \\ 1 & 4 \end{bmatrix} - \begin{bmatrix} 4 & 1 \\ 3 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -2 \\ 2 & 6 \end{bmatrix}$$

Contributed by Chris Black Solution [455]

C14 Find α and β that solve the following equation:

$$\alpha \begin{bmatrix} 1 & 2 \\ 4 & 1 \end{bmatrix} + \beta \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 4 \\ 6 & 1 \end{bmatrix}$$

Contributed by Chris Black Solution [456]

In Chapter V [197] we defined the operations of vector addition and vector scalar multiplication in Definition CVA [201] and Definition CVSM [203]. These two operations formed the underpinnings of the remainder of the chapter. We have now defined similar operations for matrices in Definition MA [427] and Definition MSM [428]. You will have noticed the resulting similarities between Theorem VSPCV [205] and Theorem VSPM [430].

In Exercises M20–M25, you will be asked to extend these similarities to other fundamental definitions and concepts we first saw in Chapter V [197]. This sequence of problems was suggested by Martin Jackson.

M20 Suppose $S = \{B_1, B_2, B_3, \ldots, B_p\}$ is a set of matrices from M_{mn} . Formulate appropriate definitions for the following terms and give an example of the use of each.

1. A linear combination of elements of S.

- 2. A relation of linear dependence on S, both trivial and non-trivial.
- 3. S is a linearly independent set.

4. $\langle S \rangle$.

Contributed by Robert Beezer

M21 Show that the set S is linearly independent in $M_{2,2}$.

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [457]

 $\mathbf{M22} \quad \mathrm{Determine \ if \ the \ set}$

$$S = \left\{ \begin{bmatrix} -2 & 3 & 4 \\ -1 & 3 & -2 \end{bmatrix}, \begin{bmatrix} 4 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -2 \\ 2 & 2 & 2 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & -2 \end{bmatrix}, \begin{bmatrix} -1 & 2 & -2 \\ 0 & -1 & -2 \end{bmatrix} \right\}$$

is linearly independent in $M_{2,3}$.

Contributed by Robert Beezer Solution [457]

M23 Determine if the matrix A is in the span of S. In other words, is $A \in \langle S \rangle$? If so write A as a linear combination of the elements of S.

$$A = \begin{bmatrix} -13 & 24 & 2\\ -8 & -2 & -20 \end{bmatrix}$$

$$S = \left\{ \begin{bmatrix} -2 & 3 & 4 \\ -1 & 3 & -2 \end{bmatrix}, \begin{bmatrix} 4 & -2 & 2 \\ 0 & -1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -2 \\ 2 & 2 & 2 \end{bmatrix}, \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & -2 \end{bmatrix}, \begin{bmatrix} -1 & 2 & -2 \\ 0 & -1 & -2 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [458]

M24 Suppose Y is the set of all 3×3 symmetric matrices (Definition SYM [434]). Find a set T so that T is linearly independent and $\langle T \rangle = Y$. Contributed by Robert Beezer Solution [459]

M25 Define a subset of $M_{3,3}$ by

$$U_{33} = \left\{ A \in M_{3,3} | [A]_{ij} = 0 \text{ whenever } i > j \right\}$$

Find a set R so that R is linearly independent and $\langle R \rangle = U_{33}$. Contributed by Robert Beezer

T13 Prove Property CM [430] of Theorem VSPM [430]. Write your proof in the style of the proof of Property DSAM [431] given in this section. Contributed by Robert Beezer Solution [459]

T14 Prove Property AAM [430] of Theorem VSPM [430]. Write your proof in the style of the proof of Property DSAM [431] given in this section. Contributed by Robert Beezer

T17 Prove Property SMAM [431] of Theorem VSPM [430]. Write your proof in the style of the proof of Property DSAM [431] given in this section.

Contributed by Robert Beezer

T18 Prove Property DMAM [431] of Theorem VSPM [430]. Write your proof in the style of the proof of Property DSAM [431] given in this section. Contributed by Robert Beezer

A matrix A is **skew-symmetric** if $A^t = -A$ Exercises T30–T37 employ this definition.

T30 Prove that a skew-symmetric matrix is square. (Hint: study the proof of Theorem SMS [435].) Contributed by Robert Beezer

T31 Prove that a skew-symmetric matrix must have zeros for its diagonal elements. In other words, if A is skew-symmetric of size n, then $[A]_{ii} = 0$ for $1 \le i \le n$. (Hint: carefully construct an example of a 3×3 skew-symmetric matrix before attempting a proof.) Contributed by Manley Perkel

T32 Prove that a matrix A is both skew-symmetric and symmetric if and only if A is the zero matrix. (Hint: one half of this proof is very easy, the other half takes a little more work.)

Contributed by Manley Perkel

T33 Suppose A and B are both skew-symmetric matrices of the same size and $\alpha, \beta \in \mathbb{C}$. Prove that $\alpha A + \beta B$ is a skew-symmetric matrix. Contributed by Manley Perkel **T34** Suppose A is a square matrix. Prove that $A + A^t$ is a symmetric matrix. Contributed by Manley Perkel

T35 Suppose A is a square matrix. Prove that $A - A^t$ is a skew-symmetric matrix. Contributed by Manley Perkel

T36 Suppose A is a square matrix. Prove that there is a symmetric matrix B and a skew-symmetric matrix C such that A = B + C. In other words, any square matrix can be decomposed into a symmetric matrix and a skew-symmetric matrix (Technique DC [1628]). (Hint: consider building a proof on Exercise MO.T34 [452] and Exercise MO.T35 [452].) Contributed by Manley Perkel

T37 Prove that the decomposition in Exercise MO.T36 [452] is unique (see Technique U [1625]). (Hint: a proof can turn on Exercise MO.T31 [451].) Contributed by Manley Perkel

Subsection SOL Solutions

C10 Contributed by Chris Black Statement [446]

1.
$$A + B = \begin{bmatrix} 4 & 6 & -2 \\ 4 & -3 & 5 \end{bmatrix}$$

2. A + C is undefined; A and C are not the same size.

3.
$$B^t + C = \begin{bmatrix} 5 & 2 \\ 6 & -6 \\ -1 & 7 \end{bmatrix}$$

4. $A + B^t$ is undefined; A and B^t are not the same size.

5.
$$\beta C = \begin{bmatrix} 1 & 2 \\ 2 & 0 \\ -1 & 1 \end{bmatrix}$$

6. $4A - 3B = \begin{bmatrix} -5 & 10 & -15 \\ 30 & 30 & -15 \end{bmatrix}$

7.
$$A^{t} + \alpha C = \begin{bmatrix} 9 & 22 \\ 20 & 3 \\ -11 & 8 \end{bmatrix}$$

8. $A + B - C^{t} = \begin{bmatrix} 2 & 2 & 0 \\ 0 & -3 & 3 \end{bmatrix}$
9. $4A + 2B - 5C^{t} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

C11 Contributed by Chris Black Statement [447] The given equation

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & 5 & -2 \end{bmatrix} = 2 \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 2 \end{bmatrix} - 3 \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & x \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 5 & 4 - 3x \end{bmatrix}$$

is valid only if 4 - 3x = -2. Thus, the only solution is x = 2.

C12 Contributed by Chris Black Statement [447] The given equation

$$\begin{bmatrix} 7 & 12 & 6\\ 6 & 4 & -2 \end{bmatrix} = \alpha \begin{bmatrix} 1 & 3 & 4\\ 2 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 4 & 3 & -6\\ 0 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} \alpha & 3\alpha & 4\alpha\\ 2\alpha & \alpha & -\alpha \end{bmatrix} + \begin{bmatrix} 4 & 3 & -6\\ 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4+\alpha & 3+3\alpha & -6+4\alpha\\ 2\alpha & 1+\alpha & 1-\alpha \end{bmatrix}$$

leads to the 6 equations in α :

$$4 + \alpha = 7$$

$$3 + 3\alpha = 12$$

$$-6 + 4\alpha = 6$$

$$2\alpha = 6$$

$$1 + \alpha = 4$$

$$1 - \alpha = -2.$$

The only value that solves all 6 equations is $\alpha = 3$, which is the solution to the original matrix equation.

C13 Contributed by Chris Black Statement [447] The given equation

$$\begin{bmatrix} 2 & 1 \\ 1 & -2 \\ 2 & 6 \end{bmatrix} = \alpha \begin{bmatrix} 3 & 1 \\ 2 & 0 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 4 & 1 \\ 3 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3\alpha - 4 & \alpha - 1 \\ 2\alpha - 3 & -2 \\ \alpha & 4\alpha - 1 \end{bmatrix}$$

gives a system of six equations in α :

$$3\alpha - 4 = 2$$

 $\alpha - 1 = 1$ $2\alpha - 3 = 1$ -2 = -2 $\alpha = 2$ $4\alpha - 1 = 6.$

Solving each of these equations, we see that the first three and the fifth all lead to the solution $\alpha = 2$, the fourth equation is true no matter what the value of α , but the last equation is only solved by $\alpha = 7/4$. Thus, the system has no solution, and the original matrix equation also has no solution.

C14 Contributed by Chris Black Statement [448] The given equation

$$\begin{bmatrix} -1 & 4\\ 6 & 1 \end{bmatrix} = \alpha \begin{bmatrix} 1 & 2\\ 4 & 1 \end{bmatrix} + \beta \begin{bmatrix} 2 & 1\\ 3 & 1 \end{bmatrix} = \begin{bmatrix} \alpha + 2\beta & 2\alpha + \beta\\ 4\alpha + 3\beta & \alpha + \beta \end{bmatrix}$$

gives a system of four equations in two variables

$$\alpha + 2\beta = -1$$
$$2\alpha + \beta = 4$$
$$4\alpha + 3\beta = 6$$
$$\alpha + \beta = 1.$$

Solving this linear system by row-reducing the augmnented matrix shows that $\alpha = 3$, $\beta = -2$ is the only solution.

M21 Contributed by Chris Black Statement [449] Suppose there exist constants α , β , γ , and δ so that

$$\alpha \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \beta \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \gamma \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \delta \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then,

$$\begin{bmatrix} \alpha & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \gamma & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so that $\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$. The only solution is then $\alpha = 0, \beta = 0, \gamma = 0$, and $\delta = 0$, so that the set S is a linearly independent set of matrices.

M22 Contributed by Chris Black Statement [449] Suppose that there exist constants a_1 , a_2 , a_3 , a_4 , and a_5 so that

$$a_{1}\begin{bmatrix} -2 & 3 & 4\\ -1 & 3 & -2 \end{bmatrix} + a_{2}\begin{bmatrix} 4 & -2 & 2\\ 0 & -1 & 1 \end{bmatrix} + a_{3}\begin{bmatrix} -1 & -2 & -2\\ 2 & 2 & 2 \end{bmatrix} + a_{4}\begin{bmatrix} -1 & 1 & 0\\ -1 & 0 & 2 \end{bmatrix} + a_{5}\begin{bmatrix} -1 & 2 & -1\\ 0 & -1 & -2 \end{bmatrix} = \begin{bmatrix} 0\\ 0 \end{bmatrix}$$

Then, we have the matrix equality (Definition ME [427])

$$\begin{bmatrix} -2a_1 + 4a_2 - a_3 - a_4 - a_5 & 3a_1 - 2a_2 - 2a_3 + a_4 + 2a_5 & 4a_1 + 2a_2 - 2a_3 - 2a_5 \\ -a_1 + 2a_3 - a_4 & 3a_1 - a_2 + 2a_3 - a_5 & -2a_1 + a_2 + 2a_3 + 2a_4 - 2a_5 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Version 2.99

which yields the linear system of equations

$$-2a_{1} + 4a_{2} - a_{3} - a_{4} - a_{5} = 0$$

$$3a_{1} - 2a_{2} - 2a_{3} + a_{4} + 2a_{5} = 0$$

$$4a_{1} + 2a_{2} - 2a_{3} - 2a_{5} = 0$$

$$-a_{1} + 2a_{3} - a_{4} = 0$$

$$3a_{1} - a_{2} + 2a_{3} - a_{5} = 0$$

$$-2a_{1} + a_{2} + 2a_{3} + 2a_{4} - 2a_{5} = 0.$$

By row-reducing the associated 6×5 homogeneous system, we see that the only solution is $a_1 = a_2 = a_3 = a_4 = a_5 = 0$, so these matrices are a linearly independent subset of $M_{2,3}$.

M23 Contributed by Chris Black Statement [449] The matrix A is in the span of S, since

$$\begin{bmatrix} -13 & 24 & 2\\ -8 & -2 & -20 \end{bmatrix} = 2 \begin{bmatrix} -2 & 3 & 4\\ -1 & 3 & -2 \end{bmatrix} - 2 \begin{bmatrix} 4 & -2 & 2\\ 0 & -1 & 1 \end{bmatrix} - 3 \begin{bmatrix} -1 & -2 & -2\\ 2 & 2 & 2 \end{bmatrix} + 4 \begin{bmatrix} -1 & 2 & -2\\ 0 & -1 & -2 \end{bmatrix}$$

Note that if we were to write a complete linear combination of all of the matrices in S, then the fourth matrix would have a zero coefficient.

M24 Contributed by Chris Black Statement [450] Since any symmetric matrix is of the form

 $\begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b & 0 \\ b & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & c \\ 0 & 0 & 0 \\ c & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & e \\ 0 & e & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & f \end{bmatrix},$

Any symmetric matrix is a linear combination of the linearly independent vectors in set T below, so that $\langle T \rangle = Y$:

$$T = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

(Something to think about: How do we know that these matrices are linearly independent?)

T13 Contributed by Robert Beezer Statement [450] For all $A, B \in M_{mn}$ and for all $1 \le i \le m, 1 \le i \le n$,

$$\begin{split} [A+B]_{ij} &= [A]_{ij} + [B]_{ij} & \text{Definition MA [427]} \\ &= [B]_{ij} + [A]_{ij} & \text{Commutativity in } \mathbb{C} \\ &= [B+A]_{ij} & \text{Definition MA [427]} \end{split}$$

With equality of each entry of the matrices A + B and B + A being equal Definition ME [427] tells us the two matrices are equal.

Section MM Matrix Multiplication

We know how to add vectors and how to multiply them by scalars. Together, these operations give us the possibility of making linear combinations. Similarly, we know how to add matrices and how to multiply matrices by scalars. In this section we mix all these ideas together and produce an operation known as "matrix multiplication." This will lead to some results that are both surprising and central. We begin with a definition of how to multiply a vector by a matrix.

Subsection MVP Matrix-Vector Product

We have repeatedly seen the importance of forming linear combinations of the columns of a matrix. As one example of this, the oft-used Theorem SLSLC [229], said that every solution to a system of linear equations gives rise to a linear combination of the column vectors of the coefficient matrix that equals the vector of constants. This theorem, and others, motivate the following central definition.

Definition MVP Matrix-Vector Product

Suppose A is an $m \times n$ matrix with columns $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n$ and \mathbf{u} is a vector of size n. Then the **matrix-vector product** of A with \mathbf{u} is the linear combination

$$A\mathbf{u} = [\mathbf{u}]_1 \mathbf{A}_1 + [\mathbf{u}]_2 \mathbf{A}_2 + [\mathbf{u}]_3 \mathbf{A}_3 + \dots + [\mathbf{u}]_n \mathbf{A}_n$$

(This definition contains Notation MVP.)

So, the matrix-vector product is yet another version of "multiplication," at least in the sense that we have yet again overloaded juxtaposition of two symbols as our notation. Remember your objects, an $m \times n$ matrix times a vector of size n will create a vector of size m. So if A is rectangular, then the size of the vector changes. With all the linear combinations we have performed so far, this computation should now seem second nature.

Example MTV A matrix times a vector Consider

$$A = \begin{bmatrix} 1 & 4 & 2 & 3 & 4 \\ -3 & 2 & 0 & 1 & -2 \\ 1 & 6 & -3 & -1 & 5 \end{bmatrix} \qquad \mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ -2 \\ 3 \\ -1 \end{bmatrix}$$

 \triangle

Then

$$A\mathbf{u} = 2\begin{bmatrix} 1\\-3\\1 \end{bmatrix} + 1\begin{bmatrix} 4\\2\\6 \end{bmatrix} + (-2)\begin{bmatrix} 2\\0\\-3 \end{bmatrix} + 3\begin{bmatrix} 3\\1\\-1 \end{bmatrix} + (-1)\begin{bmatrix} 4\\-2\\5 \end{bmatrix} = \begin{bmatrix} 7\\1\\6 \end{bmatrix}.$$

We can now represent systems of linear equations compactly with a matrix-vector product (Definition MVP [461]) and column vector equality (Definition CVE [199]). This finally yields a very popular alternative to our unconventional $\mathcal{LS}(A, \mathbf{b})$ notation.

Theorem SLEMM

Systems of Linear Equations as Matrix Multiplication

The set of solutions to the linear system $\mathcal{LS}(A, \mathbf{b})$ equals the set of solutions for \mathbf{x} in the vector equation $A\mathbf{x} = \mathbf{b}$.

Proof This theorem says that two sets (of solutions) are equal. So we need to show that one set of solutions is a subset of the other, and vice versa (Definition SE [1604]). Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n$ be the columns of A. Both of these set inclusions then follow from the following chain of equivalences (Technique E [1619]),

 $\mathbf{x} \text{ is a solution to } \mathcal{LS}(A, \mathbf{b}) \\ \iff [\mathbf{x}]_1 \mathbf{A}_1 + [\mathbf{x}]_2 \mathbf{A}_2 + [\mathbf{x}]_3 \mathbf{A}_3 + \dots + [\mathbf{x}]_n \mathbf{A}_n = \mathbf{b} \qquad \text{Theorem SLSLC [229]} \\ \iff \mathbf{x} \text{ is a solution to } A\mathbf{x} = \mathbf{b} \qquad \text{Definition MVP [461]}$

 \boxtimes

Example MNSLE Matrix notation for systems of linear equations

Consider the system of linear equations from Example NSLE [56].

$$2x_1 + 4x_2 - 3x_3 + 5x_4 + x_5 = 9$$

$$3x_1 + x_2 + x_4 - 3x_5 = 0$$

$$-2x_1 + 7x_2 - 5x_3 + 2x_4 + 2x_5 = -3$$

has coefficient matrix

$$A = \begin{bmatrix} 2 & 4 & -3 & 5 & 1 \\ 3 & 1 & 0 & 1 & -3 \\ -2 & 7 & -5 & 2 & 2 \end{bmatrix}$$

and vector of constants

$$\mathbf{b} = \begin{bmatrix} 9\\0\\-3 \end{bmatrix}$$

and so will be described compactly by the vector equation $A\mathbf{x} = \mathbf{b}$.

The matrix-vector product is a very natural computation. We have motivated it by its connections with systems of equations, but here is another example.

 \boxtimes

Example MBC Money's best cities

Every year Money magazine selects several cities in the United States as the "best" cities to live in, based on a wide array of statistics about each city. This is an example of how the editors of Money might arrive at a single number that consolidates the statistics about a city. We will analyze Los Angeles, Chicago and New York City, based on four criteria: average high temperature in July (Farenheit), number of colleges and universities in a 30-mile radius, number of toxic waste sites in the Superfund environmental clean-up program and a personal crime index based on FBI statistics (average = 100, smaller is safer). It should be apparent how to generalize the example to a greater number of cities and a greater number of statistics.

We begin by building a table of statistics. The rows will be labeled with the cities, and the columns with statistical categories. These values are from *Money*'s website in early 2005.

City	Temp	Colleges	Superfund	Crime
Los Angeles	77	28	93	254
Chicago	84	38	85	363
New York	84	99	1	193

Conceivably these data might reside in a spreadsheet. Now we must combine the statistics for each city. We could accomplish this by weighting each category, scaling the values and summing them. The sizes of the weights would depend upon the numerical size of each statistic generally, but more importantly, they would reflect the editors opinions or beliefs about which statistics were most important to their readers. Is the crime index more important than the number of colleges and universities? Of course, there is no right answer to this question.

Suppose the editors finally decide on the following weights to employ: temperature, 0.23; colleges, 0.46; Superfund, -0.05; crime, -0.20. Notice how negative weights are used for undesirable statistics. Then, for example, the editors would compute for Los Angeles,

$$(0.23)(77) + (0.46)(28) + (-0.05)(93) + (-0.20)(254) = -24.86$$

This computation might remind you of an inner product, but we will produce the computations for all of the cities as a matrix-vector product. Write the table of raw statistics as a matrix

$$T = \begin{bmatrix} 77 & 28 & 93 & 254 \\ 84 & 38 & 85 & 363 \\ 84 & 99 & 1 & 193 \end{bmatrix}$$

and the weights as a vector

$$\mathbf{w} = \begin{bmatrix} 0.23 \\ 0.46 \\ -0.05 \\ -0.20 \end{bmatrix}$$

then the matrix-vector product (Definition MVP [461]) yields

$$T\mathbf{w} = (0.23) \begin{bmatrix} 77\\84\\84 \end{bmatrix} + (0.46) \begin{bmatrix} 28\\38\\99 \end{bmatrix} + (-0.05) \begin{bmatrix} 93\\85\\1 \end{bmatrix} + (-0.20) \begin{bmatrix} 254\\363\\193 \end{bmatrix} = \begin{bmatrix} -24.86\\-40.05\\26.21 \end{bmatrix}$$

This vector contains a single number for each of the cities being studied, so the editors would rank New York best (26.21), Los Angeles next (-24.86), and Chicago third (-40.05). Of course, the mayor's offices in Chicago and Los Angeles are free to counter with a different set of weights that cause their city to be ranked best. These alternative weights would be chosen to play to each cities' strengths, and minimize their problem areas.

If a speadsheet were used to make these computations, a row of weights would be entered somewhere near the table of data and the formulas in the spreadsheet would effect a matrixvector product. This example is meant to illustrate how "linear" computations (addition, multiplication) can be organized as a matrix-vector product.

Another example would be the matrix of numerical scores on examinations and exercises for students in a class. The rows would correspond to students and the columns to exams and assignments. The instructor could then assign weights to the different exams and assignments, and via a matrix-vector product, compute a single score for each student. \boxtimes

Later (much later) we will need the following theorem, which is really a technical lemma (see Technique LC [1634]). Since we are in a position to prove it now, we will. But you can safely skip it for the moment, if you promise to come back later to study the proof when the theorem is employed. At that point you will also be able to understand the comments in the paragraph following the proof.

Theorem EMMVP Equal Matrices and Matrix-Vector Products Suppose that A and B are $m \times n$ matrices such that $A\mathbf{x} = B\mathbf{x}$ for every $\mathbf{x} \in \mathbb{C}^n$. Then

A = B.

Proof We are assuming $A\mathbf{x} = B\mathbf{x}$ for all $\mathbf{x} \in \mathbb{C}^n$, so we can employ this equality for any choice of the vector \mathbf{x} . However, we'll limit our use of this equality to the standard unit vectors, \mathbf{e}_j , $1 \leq j \leq n$ (Definition SUV [409]). For all $1 \leq j \leq n$, $1 \leq i \leq m$,

$$\begin{split} [A]_{ij} &= 0 \, [A]_{i1} + \dots + 0 \, [A]_{i,j-1} + 1 \, [A]_{ij} + 0 \, [A]_{i,j+1} + \dots + 0 \, [A]_{in} \\ &= [A]_{i1} \, [\mathbf{e}_j]_1 + [A]_{i2} \, [\mathbf{e}_j]_2 + [A]_{i3} \, [\mathbf{e}_j]_3 + \dots + [A]_{in} \, [\mathbf{e}_j]_n & \text{Definition SUV [409]} \\ &= [A\mathbf{e}_j]_i & \text{Definition MVP [461]} \\ &= [B\mathbf{e}_j]_i & \text{Definition CVE [199]} \\ &= [B]_{i1} \, [\mathbf{e}_j]_1 + [B]_{i2} \, [\mathbf{e}_j]_2 + [B]_{i3} \, [\mathbf{e}_j]_3 + \dots + [B]_{in} \, [\mathbf{e}_j]_n & \text{Definition MVP [461]} \\ &= 0 \, [B]_{i1} + \dots + 0 \, [B]_{i,j-1} + 1 \, [B]_{ij} + 0 \, [B]_{i,j+1} + \dots + 0 \, [B]_{in} & \text{Definition SUV [409]} \\ &= [B]_{ij} \end{split}$$

So by Definition ME [427] the matrices A and B are equal, as desired.

You might notice from studying the proof that the hypotheses of this theorem could be "weakened" (i.e. made less restrictive). We need only suppose the equality of the matrixvector products for just the standard unit vectors (Definition SUV [409]) or any other spanning set (Definition TSVS [746]) of \mathbb{C}^n (Exercise LISS.T40 [762]). However, in practice, when we apply this theorem the stronger hypothesis will be in effect so this version of the theorem will suffice for our purposes. (If we changed the statement of the theorem to have the less restrictive hypothesis, then we would call the theorem "stronger.")

 \Box

Subsection MM Matrix Multiplication

We now define how to multiply two matrices together. Stop for a minute and think about how you might define this new operation.

Many books would present this definition much earlier in the course. However, we have taken great care to delay it as long as possible and to present as many ideas as practical based mostly on the notion of linear combinations. Towards the conclusion of the course, or when you perhaps take a second course in linear algebra, you may be in a position to appreciate the reasons for this. For now, understand that matrix multiplication is a central definition and perhaps you will appreciate its importance more by having saved it for later.

Definition MM Matrix Multiplication

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix with columns $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \ldots, \mathbf{B}_p$. Then the **matrix product** of A with B is the $m \times p$ matrix where column i is the matrix-vector product $A\mathbf{B}_i$. Symbolically,

$$AB = A \left[\mathbf{B}_1 | \mathbf{B}_2 | \mathbf{B}_3 | \dots | \mathbf{B}_p \right] = \left[A \mathbf{B}_1 | A \mathbf{B}_2 | A \mathbf{B}_3 | \dots | A \mathbf{B}_p \right].$$

 \triangle

Example PTM Product of two matrices Set

Then

$$A = \begin{bmatrix} 1 & 2 & -1 & 4 & 6 \\ 0 & -4 & 1 & 2 & 3 \\ -5 & 1 & 2 & -3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 6 & 2 & 1 \\ -1 & 4 & 3 & 2 \\ 1 & 1 & 2 & 3 \\ 6 & 4 & -1 & 2 \\ 1 & -2 & 3 & 0 \end{bmatrix}$$
$$AB = \begin{bmatrix} A \begin{bmatrix} 1 \\ -1 \\ 1 \\ 6 \\ 1 \end{bmatrix} \begin{vmatrix} A \begin{bmatrix} 6 \\ 4 \\ 1 \\ 4 \\ -2 \end{vmatrix} \begin{vmatrix} A \begin{bmatrix} 2 \\ 3 \\ 2 \\ -1 \\ 3 \end{vmatrix} \begin{vmatrix} A \begin{bmatrix} 2 \\ 3 \\ 2 \\ -1 \\ 3 \end{vmatrix} \begin{vmatrix} A \begin{bmatrix} 2 \\ 3 \\ 2 \\ 0 \end{bmatrix} \end{vmatrix} = \begin{bmatrix} 28 & 17 & 20 & 10 \\ 20 & -13 & -3 & -1 \\ -18 & -44 & 12 & -3 \end{bmatrix}$$

 \boxtimes

Is this the definition of matrix multiplication you expected? Perhaps our previous operations for matrices caused you to think that we might multiply two matrices of the *same* size, *entry-by-entry*? Notice that our current definition uses matrices of different sizes (though the number of columns in the first must equal the number of rows in the second), and the result is of a third size. Notice too in the previous example that we cannot even consider the product *BA*, since the sizes of the two matrices in this order aren't right.

But it gets weirder than that. Many of your old ideas about "multiplication" won't apply to matrix multiplication, but some still will. So make no assumptions, and don't do anything until you have a theorem that says you can. Even if the sizes are right, matrix multiplication is not commutative — order matters.

Example MMNC Matrix multiplication is not commutative Set

$$A = \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 4 & 0 \\ 5 & 1 \end{bmatrix}.$$

Then we have two square, 2×2 matrices, so Definition MM [468] allows us to multiply them in either order. We find

$$AB = \begin{bmatrix} 19 & 3\\ 6 & 2 \end{bmatrix} \qquad \qquad BA = \begin{bmatrix} 4 & 12\\ 4 & 17 \end{bmatrix}$$

and $AB \neq BA$. Not even close. It should not be hard for you to construct other pairs of matrices that do not commute (try a couple of 3×3 's). Can you find a pair of non-identical matrices that do commute?

Matrix multiplication is fundamental, so it is a natural procedure for any computational device. See: Computation MM.MMA [1575]

Subsection MMEE Matrix Multiplication, Entry-by-Entry

While certain "natural" properties of multiplication don't hold, many more do. In the next subsection, we'll state and prove the relevant theorems. But first, we need a theorem that provides an alternate means of multiplying two matrices. In many texts, this would be given as the *definition* of matrix multiplication. We prefer to turn it around and have the following formula as a consequence of our definition. It will prove useful for proofs of matrix equality, where we need to examine products of matrices, entry-by-entry.

Theorem EMP Entries of Matrix Products

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Then for $1 \le i \le m, 1 \le j \le p$, the individual entries of AB are given by

$$[AB]_{ij} = [A]_{i1} [B]_{1j} + [A]_{i2} [B]_{2j} + [A]_{i3} [B]_{3j} + \dots + [A]_{in} [B]_{nj}$$
$$= \sum_{k=1}^{n} [A]_{ik} [B]_{kj}$$

Proof Denote the columns of A as the vectors $A_1, A_2, A_3, \ldots, A_n$ and the columns of B

 \square

-

-

as the vectors $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \ldots, \mathbf{B}_p$. Then for $1 \leq i \leq m, 1 \leq j \leq p$,

$$\begin{split} [AB]_{ij} &= [A\mathbf{B}_j]_i & \text{Definition MM [468]} \\ &= \left[[\mathbf{B}_j]_1 \mathbf{A}_1 + [\mathbf{B}_j]_2 \mathbf{A}_2 + [\mathbf{B}_j]_3 \mathbf{A}_3 + \dots + [\mathbf{B}_j]_n \mathbf{A}_n \right]_i & \text{Definition MVP [461]} \\ &= \left[[\mathbf{B}_j]_1 \mathbf{A}_1 \right]_i + \left[[\mathbf{B}_j]_2 \mathbf{A}_2 \right]_i + \left[[\mathbf{B}_j]_3 \mathbf{A}_3 \right]_i + \dots + \left[[\mathbf{B}_j]_n \mathbf{A}_n \right]_i & \text{Definition CVA [201]} \\ &= [\mathbf{B}_j]_1 [\mathbf{A}_1]_i + [\mathbf{B}_j]_2 [\mathbf{A}_2]_i + [\mathbf{B}_j]_3 [\mathbf{A}_3]_i + \dots + [\mathbf{B}_j]_n [\mathbf{A}_n]_i & \text{Definition CVSM [203]} \\ &= [B]_{1j} [A]_{i1} + [B]_{2j} [A]_{i2} + [B]_{3j} [A]_{i3} + \dots + [B]_{nj} [A]_{in} & \text{Notation ME [427]} \\ &= [A]_{i1} [B]_{1j} + [A]_{i2} [B]_{2j} + [A]_{i3} [B]_{3j} + \dots + [A]_{in} [B]_{nj} & \text{Property CMCN [1597]} \\ &= \sum_{k=1}^n [A]_{ik} [B]_{kj} \end{split}$$

Example PTMEE Product of two matrices, entry-by-entry

Consider again the two matrices from Example PTM [468]

$$A = \begin{bmatrix} 1 & 2 & -1 & 4 & 6 \\ 0 & -4 & 1 & 2 & 3 \\ -5 & 1 & 2 & -3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 6 & 2 & 1 \\ -1 & 4 & 3 & 2 \\ 1 & 1 & 2 & 3 \\ 6 & 4 & -1 & 2 \\ 1 & -2 & 3 & 0 \end{bmatrix}$$

Then suppose we just wanted the entry of AB in the second row, third column:

$$[AB]_{23} = [A]_{21} [B]_{13} + [A]_{22} [B]_{23} + [A]_{23} [B]_{33} + [A]_{24} [B]_{43} + [A]_{25} [B]_{53}$$

=(0)(2) + (-4)(3) + (1)(2) + (2)(-1) + (3)(3) = -3

Notice how there are 5 terms in the sum, since 5 is the common dimension of the two matrices (column count for A, row count for B). In the conclusion of Theorem EMP [471], it would be the index k that would run from 1 to 5 in this computation. Here's a bit more practice.

The entry of third row, first column:

$$[AB]_{31} = [A]_{31} [B]_{11} + [A]_{32} [B]_{21} + [A]_{33} [B]_{31} + [A]_{34} [B]_{41} + [A]_{35} [B]_{51}$$

= (-5)(1) + (1)(-1) + (2)(1) + (-3)(6) + (4)(1) = -18

To get some more practice on your own, complete the computation of the other 10 entries of this product. Construct some other pairs of matrices (of compatible sizes) and compute their product two ways. First use Definition MM [468]. Since linear combinations are straightforward for you now, this should be easy to do and to do correctly. Then do it again, using Theorem EMP [471]. Since this process may take some practice, use your first computation to check your work.

Theorem EMP [471] is the way many people compute matrix products by hand. It will also be very useful for the theorems we are going to prove shortly. However, the definition (Definition MM [468]) is frequently the most useful for its connections with deeper ideas like the null space and the upcoming column space.

Subsection PMM Properties of Matrix Multiplication

In this subsection, we collect properties of matrix multiplication and its interaction with the zero matrix (Definition ZM [433]), the identity matrix (Definition IM [175]), matrix addition (Definition MA [427]), scalar matrix multiplication (Definition MSM [428]), the inner product (Definition IP [398]), conjugation (Theorem MMCC [481]), and the transpose (Definition TM [433]). Whew! Here we go. These are great proofs to practice with, so try to concoct the proofs before reading them, they'll get progressively more complicated as we go.

Theorem MMZM Matrix Multiplication and the Zero Matrix

Suppose A is an $m \times n$ matrix. Then

1.
$$A\mathcal{O}_{n \times p} = \mathcal{O}_{m \times p}$$

2. $\mathcal{O}_{p \times m} A = \mathcal{O}_{p \times n}$

Proof We'll prove (1) and leave (2) to you. Entry-by-entry, for $1 \le i \le m, 1 \le j \le p$,

$$\left[A\mathcal{O}_{n\times p}\right]_{ij} = \sum_{k=1}^{n} \left[A\right]_{ik} \left[\mathcal{O}_{n\times p}\right]_{kj}$$

Theorem EMP [471]

$$= \sum_{k=1}^{n} [A]_{ik} 0$$
 Definition ZM [433]
$$= \sum_{k=1}^{n} 0$$

$$= 0$$
 Property ZCN [1597]
$$= [\mathcal{O}_{m \times p}]_{ij}$$
 Definition ZM [433]

So by the definition of matrix equality (Definition ME [427]), the matrices $A\mathcal{O}_{n\times p}$ and $\mathcal{O}_{m\times p}$ are equal.

Theorem MMIM Matrix Multiplication and Identity Matrix

Suppose A is an $m \times n$ matrix. Then

1.
$$AI_n = A$$

$$2. \quad I_m A = A$$

 \square

Proof Again, we'll prove (1) and leave (2) to you. Entry-by-entry, For $1 \le i \le m$, $1 \le j \le n$,

$$[AI_n]_{ij} = \sum_{k=1}^n [A]_{ik} [I_n]_{kj}$$
 Theorem EMP [471]

$$= [A]_{ij} [I_n]_{jj} + \sum_{\substack{k=1 \ k \neq j}}^n [A]_{ik} [I_n]_{kj}$$
Property CACN [1597]
$$= [A]_{ij} (1) + \sum_{\substack{k=1, k \neq j}}^n [A]_{ik} (0)$$
Definition IM [175]
$$= [A]_{ij} + \sum_{\substack{k=1, k \neq j}}^n 0$$
$$= [A]_{ij}$$

So the matrices A and AI_n are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

It is this theorem that gives the identity matrix its name. It is a matrix that behaves with matrix multiplication like the scalar 1 does with scalar multiplication. To multiply by the identity matrix is to have no effect on the other matrix.

Theorem MMDAA

Matrix Multiplication Distributes Across Addition

Suppose A is an $m \times n$ matrix and B and C are $n \times p$ matrices and D is a $p \times s$ matrix. Then

1.
$$A(B+C) = AB + AC$$

$$2. \quad (B+C)D = BD + CD$$

Proof We'll do (1), you do (2). Entry-by-entry, for $1 \le i \le m, 1 \le j \le p$,

$$[A(B+C)]_{ij} = \sum_{k=1}^{n} [A]_{ik} [B+C]_{kj}$$
 Theorem EMP [471]

$$= \sum_{k=1}^{n} [A]_{ik} ([B]_{kj} + [C]_{kj})$$
 Definition MA [427]

$$= \sum_{k=1}^{n} [A]_{ik} [B]_{kj} + [A]_{ik} [C]_{kj}$$
 Property DCN [1597]

$$= \sum_{k=1}^{n} [A]_{ik} [B]_{kj} + \sum_{k=1}^{n} [A]_{ik} [C]_{kj}$$
 Property CACN [1597]

$$= [AB]_{ij} + [AC]_{ij}$$
 Theorem EMP [471]

$$= [AB + AC]_{ij}$$
 Definition MA [427]

So the matrices A(B + C) and AB + AC are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

Theorem MMSMM Matrix Multiplication and Scalar Matrix Multiplication Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Let α be a scalar. Then $\alpha(AB) =$

 \square

 $(\alpha A)B = A(\alpha B).$

Proof These are equalities of matrices. We'll do the first one, the second is similar and will be good practice for you. For $1 \le i \le m$, $1 \le j \le p$,

$$\begin{aligned} [\alpha(AB)]_{ij} &= \alpha [AB]_{ij} & \text{Definition MSM [428]} \\ &= \alpha \sum_{k=1}^{n} [A]_{ik} [B]_{kj} & \text{Theorem EMP [471]} \\ &= \sum_{k=1}^{n} \alpha [A]_{ik} [B]_{kj} & \text{Property DCN [1597]} \\ &= \sum_{k=1}^{n} [\alpha A]_{ik} [B]_{kj} & \text{Definition MSM [428]} \\ &= [(\alpha A)B]_{ij} & \text{Theorem EMP [471]} \end{aligned}$$

So the matrices $\alpha(AB)$ and $(\alpha A)B$ are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

Theorem MMA

Matrix Multiplication is Associative

Suppose A is an $m \times n$ matrix, B is an $n \times p$ matrix and D is a $p \times s$ matrix. Then A(BD) = (AB)D.

Proof A matrix equality, so we'll go entry-by-entry, no surprise there. For $1 \le i \le m$,

Π

 $1\leq j\leq s,$

$$[A(BD)]_{ij} = \sum_{k=1}^{n} [A]_{ik} [BD]_{kj}$$
 Theorem EMP [471]
$$= \sum_{k=1}^{n} [A]_{ik} \left(\sum_{\ell=1}^{p} [B]_{k\ell} [D]_{\ell j} \right)$$
 Theorem EMP [471]
$$= \sum_{k=1}^{n} \sum_{\ell=1}^{p} [A]_{ik} [B]_{k\ell} [D]_{\ell j}$$
 Property DCN [1597]

We can switch the order of the summation since these are finite sums,

$$= \sum_{\ell=1}^{p} \sum_{k=1}^{n} [A]_{ik} [B]_{k\ell} [D]_{\ell j} \qquad \text{Property CACN [1597]}$$

As $[D]_{\ell j}$ does not depend on the index k, we can use distributivity to move it outside of the inner sum,

$$= \sum_{\ell=1}^{p} [D]_{\ell j} \left(\sum_{k=1}^{n} [A]_{ik} [B]_{k\ell} \right)$$
Property DCN [1597]
$$= \sum_{\ell=1}^{p} [D]_{\ell j} [AB]_{i\ell}$$
Theorem EMP [471]

$$= \sum_{\ell=1}^{p} [AB]_{i\ell} [D]_{\ell j}$$
Property CMCN [1597]
$$= [(AB)D]_{ij}$$
Theorem EMP [471]

So the matrices (AB)D and A(BD) are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

The statement of our next theorem is technically inaccurate. If we upgrade the vectors \mathbf{u}, \mathbf{v} to matrices with a single column, then the expression $\mathbf{u}^t \overline{\mathbf{v}}$ is a 1×1 matrix, though we will treat this small matrix as if it was simply the scalar quantity in its lone entry. When we apply Theorem MMIP [480] there should not be any confusion. Notice that if we treat a column vector as a matrix with a single column, then we can also construct the adjoint of a vector, though we will not make this a common practice.

Theorem MMIP

Matrix Multiplication and Inner Products

If we consider the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^m$ as $m \times 1$ matrices then

$$\langle \mathbf{u}, \, \mathbf{v}
angle = \overline{\mathbf{u}}^t \mathbf{v} = \mathbf{u}^* \mathbf{v}$$

Proof

$$\langle \mathbf{u}, \, \mathbf{v}
angle = \sum_{k=1}^m \overline{[\mathbf{u}]_k} \, [\mathbf{v}]_k$$

Definition IP [398]

$$= \sum_{k=1}^{m} \overline{[\mathbf{u}]_{k1}} [\mathbf{v}]_{k1}$$
Column vectors as matrices
$$= \sum_{k=1}^{m} \overline{[\mathbf{u}]_{k1}} [\mathbf{v}]_{k1}$$
Definition CCM [438]
$$= \sum_{k=1}^{m} \overline{[\mathbf{u}^{t}]_{1k}} [\mathbf{v}]_{k1}$$
Definition TM [433]
$$= \overline{[\mathbf{u}^{t}\mathbf{v}]_{11}}$$
Theorem EMP [471]

To finish we just blur the distinction between a 1×1 matrix $(\overline{\mathbf{u}}^t \mathbf{v})$ and its lone entry.

Theorem MMCC Matrix Multiplication and Complex Conjugation

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Then $\overline{AB} = \overline{A} \overline{B}$.

Proof To obtain this matrix equality, we will work entry-by-entry. For $1 \le i \le m$, $1 \le j \le p$,

$$\overline{AB}]_{ij} = \overline{[AB]_{ij}}$$
$$= \overline{\sum_{k=1}^{n} [A]_{ik} [B]_{kj}}$$

Definition CCM [438]

Theorem EMP [471]

$$= \sum_{k=1}^{n} \overline{[A]_{ik} [B]_{kj}}$$
Theorem CCRA [1599]
$$= \sum_{k=1}^{n} \overline{[A]_{ik}} \overline{[B]_{kj}}$$
Theorem CCRM [1599]
$$= \sum_{k=1}^{n} \overline{[A]_{ik} [B]_{kj}}$$
Definition CCM [438]
$$= \overline{[A \overline{B}]_{ij}}$$
Theorem EMP [471]

So the matrices \overline{AB} and \overline{AB} are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

Another theorem in this style, and it's a good one. If you've been practicing with the previous proofs you should be able to do this one yourself.

Theorem MMT Matrix Multiplication and Transposes

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Then $(AB)^t = B^t A^t$.

Proof This theorem may be surprising but if we check the sizes of the matrices involved, then maybe it will not seem so far-fetched. First, AB has size $m \times p$, so its transpose has size $p \times m$. The product of B^t with A^t is a $p \times n$ matrix times an $n \times m$ matrix, also resulting in a $p \times m$ matrix. So at least our objects are compatible for equality (and would not be, in general, if we didn't reverse the order of the matrix multiplication).

Here we go again, entry-by-entry. For $1 \le i \le m, 1 \le j \le p$,

$$\begin{aligned} (AB)^{t}]_{ji} &= [AB]_{ij} & \text{Definition TM [433]} \\ &= \sum_{k=1}^{n} [A]_{ik} [B]_{kj} & \text{Theorem EMP [471]} \\ &= \sum_{k=1}^{n} [B]_{kj} [A]_{ik} & \text{Property CMCN [1597]} \\ &= \sum_{k=1}^{n} [B^{t}]_{jk} [A^{t}]_{ki} & \text{Definition TM [433]} \\ &= [B^{t}A^{t}]_{ji} & \text{Theorem EMP [471]} \end{aligned}$$

So the matrices $(AB)^t$ and B^tA^t are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

This theorem seems odd at first glance, since we have to switch the order of A and B. But if we simply consider the sizes of the matrices involved, we can see that the switch is necessary for this reason alone. That the individual entries of the products then come along to be equal is a bonus.

As the adjoint of a matrix is a composition of a conjugate and a transpose, its interaction with matrix multiplication is similar to that of a transpose. Here's the last of our long list of basic properties of matrix multiplication.

Theorem MMAD Matrix Multiplication and Adjoints

Suppose A is an $m \times n$ matrix and B is an $n \times p$ matrix. Then $(AB)^* = B^*A^*$.

Proof

$(AB)^* = \left(\overline{AB}\right)^t$	Definition A [442]	
$=\left(\overline{A}\overline{B} ight)^t$	Theorem MMCC [481]	
$=\left(\overline{B} ight)^{t}\left(\overline{A} ight)^{t}$	Theorem MMT [482]	
$= B^*A^*$	Definition A [442]	

Notice how none of these proofs above relied on writing out huge general matrices with lots of ellipses ("...") and trying to formulate the equalities a whole matrix at a time. This messy business is a "proof technique" to be avoided at all costs. Notice too how the proof of Theorem MMAD [483] does not use an entry-by-entry approach, but simply builds on previous results about matrix multiplication's interaction with conjugation and transposes.

These theorems, along with Theorem VSPM [430] and the other results in Section MO [426], give you the "rules" for how matrices interact with the various operations we have defined on matrices (addition, scalar multiplication, matrix multiplication, conjugation, transposes and adjoints). Use them and use them often. But don't try to do anything with a matrix that you don't have a rule for. Together, we would informally call all these operations, and the attendant theorems, "the algebra of matrices." Notice, too, that every column

vector is just a $n \times 1$ matrix, so these theorems apply to column vectors also. Finally, these results, taken as a whole, may make us feel that the definition of matrix multiplication is not so unnatural.

Subsection HM Hermitian Matrices

The adjoint of a matrix has a basic property when employed in a matrix-vector product as part of an inner product. At this point, you could even use the following result as a motivation for the definition of an adjoint.

Theorem AIP Adjoint and Inner Product Suppose that A is an $m \times n$ matrix and $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{y} \in \mathbb{C}^m$. Then $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^*\mathbf{y} \rangle$. \Box

Proof

$$A\mathbf{x}, \mathbf{y} \rangle = \left(\overline{A\mathbf{x}}\right)^{t} \mathbf{y}$$
$$= \left(\overline{A} \,\overline{\mathbf{x}}\right)^{t} \mathbf{y}$$
$$= \overline{\mathbf{x}}^{t} \overline{A}^{t} \mathbf{y}$$
$$= \overline{\mathbf{x}}^{t} \left(A^{*} \mathbf{y}\right)$$

Theorem MMIP [480] Theorem MMCC [481] Theorem MMT [482] Definition A [442] $= \langle \mathbf{x}, A^* \mathbf{y} \rangle$ Theorem MMIP [480]

Sometimes a matrix is equal to its adjoint (Definition A [442]), and these matrices have interesting properties. One of the most common situations where this occurs is when a matrix has only real number entries. Then we are simply talking about symmetric matrices (Definition SYM [434]), so you can view this as a generalization of a symmetric matrix.

Definition HM Hermitian Matrix

The square matrix A is **Hermitian** (or **self-adjoint**) if $A = A^*$.

Again, the set of real matrices that are Hermitian is exactly the set of symmetric matrices. In Section PEE [1001] we will uncover some amazing properties of Hermitian matrices, so when you get there, run back here to remind yourself of this definition. Further properties will also appear in various sections of the Topics (Part T [1829]). Right now we prove a fundamental result about Hermitian matrices, matrix vector products and inner products. As a characterization, this could be employed as a definition of a Hermitian matrix and some authors take this approach.

Theorem HMIP

Hermitian Matrices and Inner Products

Suppose that A is a square matrix of size n. Then A is Hermitian if and only if $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

Proof (\Rightarrow) This is the "easy half" of the proof, and makes the rationale for a definition

 \triangle

of Hermitian matrices most obvious. Assume A is Hermitian,

$$\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^* \mathbf{y} \rangle$$
 Theorem AIP [485]
= $\langle \mathbf{x}, A\mathbf{y} \rangle$ Definition HM [486]

(\Leftarrow) This "half" will take a bit more work. Assume that $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A\mathbf{y} \rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Choose any $\mathbf{x} \in \mathbb{C}^n$. We want to show that $A = A^*$ by establishing that $A\mathbf{x} = A^*\mathbf{x}$. With only this much motivation, consider the inner product,

$$\langle A\mathbf{x} - A^*\mathbf{x}, A\mathbf{x} - A^*\mathbf{x} \rangle = \langle A\mathbf{x} - A^*\mathbf{x}, A\mathbf{x} \rangle - \langle A\mathbf{x} - A^*\mathbf{x}, A^*\mathbf{x} \rangle$$
 Theorem IPVA [400]
$$= \langle A\mathbf{x} - A^*\mathbf{x}, A\mathbf{x} \rangle - \langle A(A\mathbf{x} - A^*\mathbf{x}), \mathbf{x} \rangle$$
 Theorem AIP [485]
$$= \langle A\mathbf{x} - A^*\mathbf{x}, A\mathbf{x} \rangle - \langle A\mathbf{x} - A^*\mathbf{x}, A\mathbf{x} \rangle$$
 Hypothesis
$$= 0$$
 Property AICN [1597]

Because this first inner product equals zero, and has the same vector in each argument $(A\mathbf{x} - A^*\mathbf{x})$, Theorem PIP [406] gives the conclusion that $A\mathbf{x} - A^*\mathbf{x} = \mathbf{0}$. With $A\mathbf{x} = A^*\mathbf{x}$ for all $\mathbf{x} \in \mathbb{C}^n$, Theorem EMMVP [466] says $A = A^*$, which is the defining property of a Hermitian matrix (Definition HM [486]).

So, informally, Hermitian matrices are those that can be tossed around from one side of an inner product to the other with reckless abandon. We'll see later what this buys us.

Subsection READ Reading Questions

1. Form the matrix vector product of

$$\begin{bmatrix} 2 & 3 & -1 & 0 \\ 1 & -2 & 7 & 3 \\ 1 & 5 & 3 & 2 \end{bmatrix} \qquad \text{with} \qquad \begin{bmatrix} 2 \\ -3 \\ 0 \\ 5 \end{bmatrix}$$

2. Multiply together the two matrices below (in the order given).

$$\begin{bmatrix} 2 & 3 & -1 & 0 \\ 1 & -2 & 7 & 3 \\ 1 & 5 & 3 & 2 \end{bmatrix} \qquad \qquad \begin{bmatrix} 2 & 6 \\ -3 & -4 \\ 0 & 2 \\ 3 & -1 \end{bmatrix}$$

3. Rewrite the system of linear equations below as a vector equality and using a matrixvector product. (This question does not ask for a solution to the system. But it does ask you to express the system of equations in a new form using tools from this section.)

$$2x_1 + 3x_2 - x_3 = 0$$

$$x_1 + 2x_2 + x_3 = 3$$
$$x_1 + 3x_2 + 3x_3 = 7$$

Subsection EXC Exercises

C20 Compute the product of the two matrices below, AB. Do this using the definitions of the matrix-vector product (Definition MVP [461]) and the definition of matrix multiplication (Definition MM [468]).

$$A = \begin{bmatrix} 2 & 5\\ -1 & 3\\ 2 & -2 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 1 & 5 & -3 & 4\\ 2 & 0 & 2 & -3 \end{bmatrix}$$

Contributed by Robert Beezer Solution [496]

C21 Compute the product AB of the two matrices below using both the definition of the matrix-vector product (Definition MVP [461]) and the definition of matrix multiplication (Definition MM [468]).

$$A = \begin{bmatrix} 1 & 3 & 2 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 0 & 1 \\ 3 & 1 & 5 \end{bmatrix}$$

Contributed by Chris Black Solution [496]

C22 Compute the product AB of the two matrices below using both the definition of the matrix-vector product (Definition MVP [461]) and the definition of matrix multiplication (Definition MM [468]).

$$A = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix}$$

Contributed by Chris Black Solution [497]

C23 Compute the product AB of the two matrices below using both the definition of the matrix-vector product (Definition MVP [461]) and the definition of matrix multiplication (Definition MM [468]).

$$A = \begin{bmatrix} 3 & 1\\ 2 & 4\\ 6 & 5\\ 1 & 2 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} -3 & 1\\ 4 & 2 \end{bmatrix}$$

Contributed by Chris Black Solution [497]

C24 Compute the product AB of the two matrices below.

$$A = \begin{bmatrix} 1 & 2 & 3 & -2 \\ 0 & 1 & -2 & -1 \\ 1 & 1 & 3 & 1 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 3 \\ 4 \\ 0 \\ 2 \end{bmatrix}$$

Contributed by Chris Black Solution [497]

C25 Compute the product AB of the two matrices below.

$$A = \begin{bmatrix} 1 & 2 & 3 & -2 \\ 0 & 1 & -2 & -1 \\ 1 & 1 & 3 & 1 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} -7 \\ 3 \\ 1 \\ 1 \end{bmatrix}$$

Contributed by Chris Black Solution [497]

C26 Compute the product AB of the two matrices below using both the definition of the matrix-vector product (Definition MVP [461]) and the definition of matrix multiplication (Definition MM [468]).

$$A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 2 & -5 & -1 \\ 0 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

Contributed by Chris Black Solution [497]

C30 For the matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, find A^2 , A^3 , A^4 . Find a general formula for A^n for any positive integer n.

Contributed by Chris Black Solution [498]

C31 For the matrix $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, find A^2 , A^3 , A^4 . Find a general formula for A^n for any positive integer n. Contributed by Chris Black Solution [498] **C32** For the matrix $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, find A^2 , A^3 , A^4 . Find a general formula for A^n for any positive integer n. Contributed by Chris Black Solution [498]

C33 For the matrix
$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$
, find A^2 , A^3 , A^4 . Find a general formula for A^n for

any positive integer n.

Contributed by Chris Black Solution [498]

T10 Suppose that A is a square matrix and there is a vector, **b**, such that $\mathcal{LS}(A, \mathbf{b})$ has a unique solution. Prove that A is nonsingular. Give a direct proof (perhaps appealing to Theorem PSPHS [256]) rather than just negating a sentence from the text discussing a similar situation.

Contributed by Robert Beezer Solution [498]

T20 Prove the second part of Theorem MMZM [474]. Contributed by Robert Beezer **T21** Prove the second part of Theorem MMIM [475]. Contributed by Robert Beezer

T22 Prove the second part of Theorem MMDAA [476]. Contributed by Robert Beezer

T23 Prove the second part of Theorem MMSMM [477]. Contributed by Robert Beezer Solution [499]

T31 Suppose that A is an $m \times n$ matrix and $\mathbf{x}, \mathbf{y} \in \mathcal{N}(A)$. Prove that $\mathbf{x} + \mathbf{y} \in \mathcal{N}(A)$. Contributed by Robert Beezer

T32 Suppose that A is an $m \times n$ matrix, $\alpha \in \mathbb{C}$, and $\mathbf{x} \in \mathcal{N}(A)$. Prove that $\alpha \mathbf{x} \in \mathcal{N}(A)$. Contributed by Robert Beezer

T40 Suppose that A is an $m \times n$ matrix and B is an $n \times p$ matrix. Prove that the null space of B is a subset of the null space of AB, that is $\mathcal{N}(B) \subseteq \mathcal{N}(AB)$. Provide an example where the opposite is false, in other words give an example where $\mathcal{N}(AB) \not\subseteq \mathcal{N}(B)$. Contributed by Robert Beezer Solution [500]

T41 Suppose that A is an $n \times n$ nonsingular matrix and B is an $n \times p$ matrix. Prove that the null space of B is equal to the null space of AB, that is $\mathcal{N}(B) = \mathcal{N}(AB)$. (Compare with Exercise MM.T40 [494].) Contributed by Robert Beezer Solution [500]

T50 Suppose **u** and **v** are any two solutions of the linear system $\mathcal{LS}(A, \mathbf{b})$. Prove that

 $\mathbf{u} - \mathbf{v}$ is an element of the null space of A, that is, $\mathbf{u} - \mathbf{v} \in \mathcal{N}(A)$. Contributed by Robert Beezer

T51 Give a new proof of Theorem PSPHS [256] replacing applications of Theorem SLSLC [229] with matrix-vector products (Theorem SLEMM [462]). Contributed by Robert Beezer Solution [501]

T52 Suppose that $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, $\mathbf{b} \in \mathbb{C}^m$ and A is an $m \times n$ matrix. If \mathbf{x}, \mathbf{y} and $\mathbf{x} + \mathbf{y}$ are each a solution to the linear system $\mathcal{LS}(A, \mathbf{b})$, what interesting can you say about \mathbf{b} ? Form an implication with the existence of the three solutions as the hypothesis and an interesting statement about $\mathcal{LS}(A, \mathbf{b})$ as the conclusion, and then give a proof. Contributed by Robert Beezer Solution [502]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [490] By Definition MM [468],

$$AB = \begin{bmatrix} 2 & 5\\ -1 & 3\\ 2 & -2 \end{bmatrix} \begin{bmatrix} 1\\ 2 \end{bmatrix} \begin{vmatrix} 2 & 5\\ -1 & 3\\ 2 & -2 \end{bmatrix} \begin{bmatrix} 5\\ 0 \end{bmatrix} \begin{vmatrix} 2 & 5\\ -1 & 3\\ 2 & -2 \end{bmatrix} \begin{bmatrix} -3\\ 2 \end{vmatrix} \begin{vmatrix} 2 & 5\\ -1 & 3\\ 2 & -2 \end{bmatrix} \begin{bmatrix} 4\\ -2 \end{bmatrix}$$

Repeated applications of Definition MVP [461] give

$$= \begin{bmatrix} 1 \begin{bmatrix} 2\\-1\\2 \end{bmatrix} + 2 \begin{bmatrix} 5\\3\\-2 \end{bmatrix} \begin{vmatrix} 5 \begin{bmatrix} 2\\-1\\2 \end{bmatrix} + 0 \begin{bmatrix} 5\\3\\-2 \end{bmatrix} \begin{vmatrix} -3 \begin{bmatrix} 2\\-1\\2 \end{bmatrix} + 2 \begin{bmatrix} 5\\3\\-2 \end{bmatrix} \begin{vmatrix} 4 \begin{bmatrix} 2\\-1\\2 \end{bmatrix} + (-3) \begin{bmatrix} 5\\3\\-2 \end{bmatrix} \end{vmatrix}$$
$$= \begin{bmatrix} 12 & 10 & 4 & -7\\5 & -5 & 9 & -13\\-2 & 10 & -10 & 14 \end{bmatrix}$$

C21 Contributed by Chris Black Statement [490] $AB = \begin{bmatrix} 13 & 3 & 15 \\ 1 & 0 & 5 \\ 1 & 0 & 1 \end{bmatrix}.$

C22 Contributed by Chris Black Statement [491]

$$AB = \begin{bmatrix} 2 & 3 \\ 0 & 0 \end{bmatrix}.$$

C23 Contributed by Chris Black Statement [491]

$$AB = \begin{bmatrix} -5 & 5\\ 10 & 10\\ 2 & 16\\ 5 & 5 \end{bmatrix}.$$

C24 Contributed by Chris Black Statement [491]

$$AB = \begin{bmatrix} 7\\2\\9 \end{bmatrix}.$$

C25 Contributed by Chris Black Statement [492]
$$AB = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

C26 Contributed by Chris Black Statement [492]

$$AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

C30 Contributed by Chris Black Statement [492]

$$A^{2} = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}, A^{3} = \begin{bmatrix} 1 & 6 \\ 0 & 1 \end{bmatrix}, A^{4} = \begin{bmatrix} 1 & 8 \\ 0 & 1 \end{bmatrix}.$$
 From this pattern, we see that $A^{n} = \begin{bmatrix} 1 & 2n \\ 0 & 1 \end{bmatrix}.$
C31 Contributed by Chris Black Statement [493]

$$A^{2} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}, A^{3} = \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}, A^{4} = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}.$$
 From this pattern, we see that $A^{n} = \begin{bmatrix} 1 & -n \\ 0 & 1 \end{bmatrix}.$
C32 Contributed by Chris Black Statement [493]

$$A^{2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}, A^{3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 27 \end{bmatrix}, \text{ and } A^{4} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 16 & 0 \\ 0 & 0 & 81 \end{bmatrix}.$$
 The pattern emerges, and we

see that
$$A^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 3^n \end{bmatrix}$$
.

C33 Contributed by Chris Black Statement [493] We quickly compute $A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and we then see that A^3 and all subsequent powers of A are the 3 × 3 zero matrix; that is, $A^n = \mathcal{O}_{3,3}$ for $n \ge 3$.

T10 Contributed by Robert Beezer Statement [493] Since $\mathcal{LS}(A, b)$ has at least one solution, we can apply Theorem PSPHS [256]. Because the solution is assumed to be unique, the null space of A must be trivial. Then Theorem NMTNS [179] implies that A is nonsingular.

The converse of this statement is a trivial application of Theorem NMUS [180]. That said, we could extend our NSMxx series of theorems with an added equivalence for nonsingularity, "Given a single vector of constants, **b**, the system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution."

T23 Contributed by Robert Beezer Statement [494] We'll run the proof entry-by-entry.

$$[\alpha(AB)]_{ij} = \alpha [AB]_{ij}$$
$$= \alpha \sum_{k=1}^{n} [A]_{ik} [B]_{kj}$$
$$= \sum_{k=1}^{n} \alpha [A]_{ik} [B]_{kj}$$
$$= \sum_{k=1}^{n} [A]_{ik} \alpha [B]_{kj}$$
$$= \sum_{k=1}^{n} [A]_{ik} [\alpha B]_{kj}$$
$$= [A(\alpha B)]_{ij}$$

Definition MSM [428]

Theorem EMP [471]

Distributivity in $\mathbb C$

Commutativity in $\mathbb C$

Definition MSM [428]

Theorem EMP [471]

So the matrices $\alpha(AB)$ and $A(\alpha B)$ are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [427]) we can say they are equal matrices.

T40 Contributed by Robert Beezer Statement [494]

To prove that one set is a subset of another, we start with an element of the smaller set and see if we can determine that it is a member of the larger set (Definition SSET [1603]). Suppose $\mathbf{x} \in \mathcal{N}(B)$. Then we know that $B\mathbf{x} = \mathbf{0}$ by Definition NSM [152]. Consider

$(AB)\mathbf{x} = A(B\mathbf{x})$	Theorem MMA $[478]$
= A 0	Hypothesis
= 0	Theorem MMZM [474]

This establishes that $\mathbf{x} \in \mathcal{N}(AB)$, so $\mathcal{N}(B) \subseteq \mathcal{N}(AB)$.

To show that the inclusion does not hold in the opposite direction, choose B to be any nonsingular matrix of size n. Then $\mathcal{N}(B) = \{\mathbf{0}\}$ by Theorem NMTNS [179]. Let A be the square zero matrix, \mathcal{O} , of the same size. Then $AB = \mathcal{O}B = \mathcal{O}$ by Theorem MMZM [474] and therefore $\mathcal{N}(AB) = \mathbb{C}^n$, and is not a subset of $\mathcal{N}(B) = \{\mathbf{0}\}$.

T41 Contributed by David Braithwaite Statement [494] From the solution to Exercise MM.T40 [494] we know that $\mathcal{N}(B) \subseteq \mathcal{N}(AB)$. So to establish the set equality (Definition SE [1604]) we need to show that $\mathcal{N}(AB) \subseteq \mathcal{N}(B)$.

Suppose $\mathbf{x} \in \mathcal{N}(AB)$. Then we know that $AB\mathbf{x} = \mathbf{0}$ by Definition NSM [152]. Consider

$0 = (AB)\mathbf{x}$	Definition NSM $[152]$
$=A\left(B\mathbf{x} ight)$	Theorem MMA [478]

So, $B\mathbf{x} \in \mathcal{N}(A)$. Because A is nonsingular, it has a trivial null space (Theorem NMTNS [179]) and we conclude that $B\mathbf{x} = \mathbf{0}$. This establishes that $\mathbf{x} \in \mathcal{N}(B)$, so $\mathcal{N}(AB) \subseteq \mathcal{N}(B)$ and combined with the solution to Exercise MM.T40 [494] we have $\mathcal{N}(B) = \mathcal{N}(AB)$ when A is nonsingular.

T51 Contributed by Robert Beezer Statement [495]

We will work with the vector equality representations of the relevant systems of equations, as described by Theorem SLEMM [462].

(\Leftarrow) Suppose $\mathbf{y} = \mathbf{w} + \mathbf{z}$ and $\mathbf{z} \in \mathcal{N}(A)$. Then

$A\mathbf{y} = A(\mathbf{w} + \mathbf{z})$	Substitution		
$= A\mathbf{w} + A\mathbf{z}$	Theorem MMDAA [476]		
$= \mathbf{b} + 0$	$\mathbf{z} \in \mathcal{N}(A)$		
= b	Property ZC [206]		

demonstrating that \mathbf{y} is a solution.

 (\Rightarrow) Suppose **y** is a solution to $\mathcal{LS}(A, b)$. Then

$$A(\mathbf{y} - \mathbf{w}) = A\mathbf{y} - A\mathbf{w}$$

Theorem MMDAA [476]
$$= \mathbf{b} - \mathbf{b}$$

$$= \mathbf{0}$$

Theorem MMDAA [476]
$$\mathbf{y}, \mathbf{w} \text{ solutions to } A\mathbf{x} = \mathbf{b}$$

Property AIC [206]

which says that $\mathbf{y} - \mathbf{w} \in \mathcal{N}(A)$. In other words, $\mathbf{y} - \mathbf{w} = \mathbf{z}$ for some vector $\mathbf{z} \in \mathcal{N}(A)$. Rewritten, this is $\mathbf{y} = \mathbf{w} + \mathbf{z}$, as desired.

T52 Contributed by Robert Beezer Statement [495] $\mathcal{LS}(A, \mathbf{b})$ must be homogeneous. To see this consider that

$\mathbf{b} = A\mathbf{x}$	Theorem SLEMM [462]
$= A\mathbf{x} + 0$	Property ZC $[206]$
$= A\mathbf{x} + A\mathbf{y} - A\mathbf{y}$	Property AIC [206]
$=A\left(\mathbf{x}+\mathbf{y}\right)-A\mathbf{y}$	Theorem MMDAA [476]
$= \mathbf{b} - \mathbf{b}$	Theorem SLEMM [462]
= 0	Property AIC [206]

By Definition HS [146] we see that $\mathcal{LS}(A, \mathbf{b})$ is homogeneous.

Section MISLE Matrix Inverses and Systems of Linear Equations

We begin with a familiar example, performed in a novel way.

Example SABMI Solutions to Archetype B with a matrix inverse Archetype B [1652] is the system of m = 3 linear equations in n = 3 variables,

$$-7x_1 - 6x_2 - 12x_3 = -33$$

$$5x_1 + 5x_2 + 7x_3 = 24$$

$$x_1 + 4x_3 = 5$$

By Theorem SLEMM [462] we can represent this system of equations as

$$A\mathbf{x} = \mathbf{b}$$

where

$$A = \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} -33 \\ 24 \\ 5 \end{bmatrix}$$

Now, entirely unmotivated, we define the 3×3 matrix B,

$$B = \begin{bmatrix} -10 & -12 & -9\\ \frac{13}{2} & 8 & \frac{11}{2}\\ \frac{5}{2} & 3 & \frac{5}{2} \end{bmatrix}$$

and note the remarkable fact that

$$BA = \begin{bmatrix} -10 & -12 & -9\\ \frac{13}{2} & 8 & \frac{11}{2}\\ \frac{5}{2} & 3 & \frac{5}{2} \end{bmatrix} \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}$$

Now apply this computation to the problem of solving the system of equations,

$\mathbf{x} = I_3 \mathbf{x}$	Theorem MMIM $[475]$	
$= (BA)\mathbf{x}$	Substitution	
$= B(A\mathbf{x})$	Theorem MMA [478]	
$= B\mathbf{b}$	Substitution	

So we have

$$\mathbf{x} = B\mathbf{b} = \begin{bmatrix} -10 & -12 & -9\\ \frac{13}{2} & 8 & \frac{11}{2}\\ \frac{5}{2} & 3 & \frac{5}{2} \end{bmatrix} \begin{bmatrix} -33\\ 24\\ 5 \end{bmatrix} = \begin{bmatrix} -3\\ 5\\ 2 \end{bmatrix}$$

So with the help and assistance of B we have been able to determine a solution to the system represented by $A\mathbf{x} = \mathbf{b}$ through judicious use of matrix multiplication. We know by Theorem NMUS [180] that since the coefficient matrix in this example is nonsingular, there would be a unique solution, no matter what the choice of \mathbf{b} . The derivation above amplifies this result, since we were *forced* to conclude that $\mathbf{x} = B\mathbf{b}$ and the solution couldn't be anything else. You should notice that this argument would hold for any particular value of \mathbf{b} .

The matrix B of the previous example is called the inverse of A. When A and B are combined via matrix multiplication, the result is the identity matrix, which can be inserted "in front" of \mathbf{x} as the first step in finding the solution. This is entirely analogous to how we might solve a single linear equation like 3x = 12.

$$x = 1x = \left(\frac{1}{3}(3)\right)x = \frac{1}{3}(3x) = \frac{1}{3}(12) = 4$$

Here we have obtained a solution by employing the "multiplicative inverse" of 3, $3^{-1} = \frac{1}{3}$. This works fine for any scalar multiple of x, except for zero, since zero does not have a multiplicative inverse. Consider separately the two linear equations,

$$0x = 12 \qquad \qquad 0x = 0$$

The first has no solutions, while the second has infinitely many solutions. For matrices, it is all just a little more complicated. Some matrices have inverses, some do not. And when a matrix does have an inverse, just how would we compute it? In other words, just where did that matrix B in the last example come from? Are there other matrices that might have worked just as well?

Subsection IM Inverse of a Matrix

Definition MI Matrix Inverse

Suppose A and B are square matrices of size n such that $AB = I_n$ and $BA = I_n$. Then A is **invertible** and B is the **inverse** of A. In this situation, we write $B = A^{-1}$. (This definition contains Notation MI.)

Notice that if B is the inverse of A, then we can just as easily say A is the inverse of B, or A and B are inverses of each other.

Not every square matrix has an inverse. In Example SABMI [503] the matrix B is the inverse of the coefficient matrix of Archetype B [1652]. To see this it only remains to check that $AB = I_3$. What about Archetype A [1642]? It is an example of a square matrix without an inverse.

Example MWIAA

A matrix without an inverse, Archetype A

Consider the coefficient matrix from Archetype A [1642],

$$A = \begin{bmatrix} 1 & -1 & 2\\ 2 & 1 & 1\\ 1 & 1 & 0 \end{bmatrix}$$

Suppose that A is invertible and does have an inverse, say B. Choose the vector of constants

$$\mathbf{b} = \begin{bmatrix} 1\\3\\2 \end{bmatrix}$$

and consider the system of equations $\mathcal{LS}(A, \mathbf{b})$. Just as in Example SABMI [503], this vector equation would have the unique solution $\mathbf{x} = B\mathbf{b}$.

However, the system $\mathcal{LS}(A, \mathbf{b})$ is inconsistent. Form the augmented matrix $[A \mid \mathbf{b}]$ and row-reduce to

$\left[1 \right]$	0	1	0]
0	1	-1	0
0	0	0	1

which allows to recognize the inconsistency by Theorem RCLS [120].

So the assumption of A's inverse leads to a logical inconsistency (the system can't be both consistent and inconsistent), so our assumption is false. A is not invertible.

It's possible this example is less than satisfying. Just where did that particular choice of the vector **b** come from anyway? Stay tuned for an application of the future Theorem CSCS [563] in Example CSAA [573].

Let's look at one more matrix inverse before we embark on a more systematic study.

Example MI Matrix inverse

Consider the matrices,

Then

and

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ -2 & -3 & 0 & -5 & -1 \\ 1 & 1 & 0 & 2 & 1 \\ -2 & -3 & -1 & -3 & -2 \\ -1 & -3 & -1 & -3 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} -3 & 3 & 6 & -1 & -2 \\ 0 & -2 & -5 & -1 & 1 \\ 1 & 2 & 4 & 1 & -1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -2 & 0 & 1 \end{bmatrix}$$
$$AB = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ -2 & -3 & 0 & -5 & -1 \\ 1 & 1 & 0 & 2 & 1 \\ -2 & -3 & -1 & -3 & -2 \\ -1 & -3 & -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} -3 & 3 & 6 & -1 & -2 \\ 0 & -2 & -5 & -1 & 1 \\ 1 & 2 & 4 & 1 & -1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
$$BA = \begin{bmatrix} -3 & 3 & 6 & -1 & -2 \\ 0 & -2 & -3 & -1 & -3 & -2 \\ 0 & -2 & -5 & -1 & 1 \\ 1 & 2 & 4 & 1 & -1 \\ 1 & 2 & 4 & 1 & -1 \\ 1 & 2 & 4 & 1 & -1 \\ 1 & 1 & 0 & 2 & 1 \\ -2 & -3 & -1 & -3 & -2 \\ -1 & -3 & -1 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

so by Definition MI [506], we can say that A is invertible and write $B = A^{-1}$.

We will now concern ourselves less with whether or not an inverse of a matrix exists, but instead with how you can find one when it does exist. In Section MINM [537] we will

 \boxtimes

have some theorems that allow us to more quickly and easily determine just when a matrix is invertible.

Subsection CIM Computing the Inverse of a Matrix

We've seen that the matrices from Archetype B [1652] and Archetype K [1734] both have inverses, but these inverse matrices have just dropped from the sky. How would we compute an inverse? And just when is a matrix invertible, and when is it not? Writing a putative inverse with n^2 unknowns and solving the resultant n^2 equations is one approach. Applying this approach to 2×2 matrices can get us somewhere, so just for fun, let's do it.

Theorem TTMI Two-by-Two Matrix Inverse Suppose

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Then A is invertible if and only if $ad - bc \neq 0$. When A is invertible, then

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Proof (\Leftarrow) Assume that $ad - bc \neq 0$. We will use the definition of the inverse of a matrix to establish that A has inverse (Definition MI [506]). Note that if $ad - bc \neq 0$ then the displayed formula for A^{-1} is legitimate since we are not dividing by zero). Using this proposed formula for the inverse of A, we compute

$$AA^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} 1 \\ ad - bc \end{bmatrix} \begin{pmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$A^{-1}A = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

By Definition MI [506] this is sufficient to establish that A is invertible, and that the expression for A^{-1} is correct.

 (\Rightarrow) Assume that A is invertible, and proceed with a proof by contradiction (Technique CD [1623]), by assuming also that ad - bc = 0. This translates to ad = bc. Let

$$B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

be a putative inverse of A. This means that

$$I_2 = AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

Working on the matrices on two ends of this equation, we will multiply the top row by c and the bottom row by a.

$$\begin{bmatrix} c & 0 \\ 0 & a \end{bmatrix} = \begin{bmatrix} ace + bcg & acf + bch \\ ace + adg & acf + adh \end{bmatrix}$$

We are assuming that ad = bc, so we can replace two occurrences of ad by bc in the bottom row of the right matrix.

$$\begin{bmatrix} c & 0 \\ 0 & a \end{bmatrix} = \begin{bmatrix} ace + bcg & acf + bch \\ ace + bcg & acf + bch \end{bmatrix}$$

The matrix on the right now has two rows that are identical, and therefore the same must be true of the matrix on the left. Identical rows for the matrix on the left implies that a = 0and c = 0.

With this information, the product AB becomes

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 = AB = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix} = \begin{bmatrix} bg & bh \\ dg & dh \end{bmatrix}$$

So bg = dh = 1 and thus b, g, d, h are all nonzero. But then bh and dg (the "other corners") must also be nonzero, so this is (finally) a contradiction. So our assumption was false and we see that $ad - bc \neq 0$ whenever A has an inverse.

There are several ways one could try to prove this theorem, but there is a continual temptation to divide by one of the eight entries involved (a through f), but we can never be sure if these numbers are zero or not. This could lead to an analysis by cases, which is messy, messy, messy. Note how the above proof never divides, but always multiplies, and

how zero/nonzero considerations are handled. Pay attention to the expression ad - bc, as we will see it again in a while (Chapter D [882]).

This theorem is cute, and it is nice to have a formula for the inverse, and a condition that tells us when we can use it. However, this approach becomes impractical for larger matrices, even though it is possible to demonstrate that, in theory, there is a general formula. (Think for a minute about extending this result to just 3×3 matrices. For starters, we need 18 letters!) Instead, we will work column-by-column. Let's first work an example that will motivate the main theorem and remove some of the previous mystery.

Example CMI Computing a matrix inverse

Consider the matrix defined in Example MI [507] as,

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 \\ -2 & -3 & 0 & -5 & -1 \\ 1 & 1 & 0 & 2 & 1 \\ -2 & -3 & -1 & -3 & -2 \\ -1 & -3 & -1 & -3 & 1 \end{bmatrix}$$

For its inverse, we desire a matrix B so that $AB = I_5$. Emphasizing the structure of the columns and employing the definition of matrix multiplication Definition MM [468],

$$AB = I_5$$
$$A[\mathbf{B}_1|\mathbf{B}_2|\mathbf{B}_3|\mathbf{B}_4|\mathbf{B}_5] = [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\mathbf{e}_4|\mathbf{e}_5]$$

$$[A\mathbf{B}_1|A\mathbf{B}_2|A\mathbf{B}_3|A\mathbf{B}_4|A\mathbf{B}_5] = [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\mathbf{e}_4|\mathbf{e}_5].$$

Equating the matrices column-by-column we have

$$A\mathbf{B}_1 = \mathbf{e}_1$$
 $A\mathbf{B}_2 = \mathbf{e}_2$ $A\mathbf{B}_3 = \mathbf{e}_3$ $A\mathbf{B}_4 = \mathbf{e}_4$ $A\mathbf{B}_5 = \mathbf{e}_5$

Since the matrix B is what we are trying to compute, we can view each column, \mathbf{B}_i , as a column vector of unknowns. Then we have five systems of equations to solve, each with 5 equations in 5 variables. Notice that all 5 of these systems have the same coefficient matrix. We'll now solve each system in turn,

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A, \mathbf{e}_1)$,

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 1 \\ -2 & -3 & 0 & -5 & -1 & 0 \\ 1 & 1 & 0 & 2 & 1 & 0 \\ -2 & -3 & -1 & -3 & -2 & 0 \\ -1 & -3 & -1 & -3 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -3 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ so } \mathbf{B}_{1} = \begin{bmatrix} -3 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A, \mathbf{e}_2)$,

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 0 \\ -2 & -3 & 0 & -5 & -1 & 1 \\ 1 & 1 & 0 & 2 & 1 & 0 \\ -2 & -3 & -1 & -3 & -2 & 0 \\ -1 & -3 & -1 & -3 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \text{ so } \mathbf{B}_2 = \begin{bmatrix} 3 \\ -2 \\ 2 \\ 0 \\ -1 \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A, \mathbf{e}_3)$,

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 0 \\ -2 & -3 & 0 & -5 & -1 & 0 \\ 1 & 1 & 0 & 2 & 1 & 1 \\ -2 & -3 & -1 & -3 & -2 & 0 \\ -1 & -3 & -1 & -3 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 & -5 \\ 0 & 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -2 \end{bmatrix} \text{ so } \mathbf{B}_{3} = \begin{bmatrix} 6 \\ -5 \\ 4 \\ 1 \\ -2 \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A, \mathbf{e}_4)$,

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 0 \\ -2 & -3 & 0 & -5 & -1 & 0 \\ 1 & 1 & 0 & 2 & 1 & 0 \\ -2 & -3 & -1 & -3 & -2 & 1 \\ -1 & -3 & -1 & -3 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \text{ so } \mathbf{B}_{4} = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A, \mathbf{e}_5)$,

$$\begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 0 \\ -2 & -3 & 0 & -5 & -1 & 0 \\ 1 & 1 & 0 & 2 & 1 & 0 \\ -2 & -3 & -1 & -3 & -2 & 0 \\ -1 & -3 & -1 & -3 & 1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ so } \mathbf{B}_{5} = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

We can now collect our 5 solution vectors into the matrix B,

$$B = [\mathbf{B}_1 | \mathbf{B}_2 | \mathbf{B}_3 | \mathbf{B}_4 | \mathbf{B}_5]$$

$$= \begin{bmatrix} \begin{bmatrix} -3\\0\\1\\1\\1\\1 \end{bmatrix} & \begin{bmatrix} 3\\-2\\2\\0\\-1 \end{bmatrix} & \begin{bmatrix} 6\\-5\\4\\1\\-1\\1\\0 \end{bmatrix} & \begin{bmatrix} -1\\-1\\1\\1\\0\\1 \end{bmatrix} & \begin{bmatrix} -2\\1\\-1\\1\\0\\1 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} -3&3&6&-1&-2\\0&-2&-5&-1&1\\1&2&4&1&-1\\1&0&1&1&0\\1&-1&-2&0&1 \end{bmatrix}$$

By this method, we know that $AB = I_5$. Check that $BA = I_5$, and then we will know that we have the inverse of A.

Notice how the five systems of equations in the preceding example were all solved by *exactly* the same sequence of row operations. Wouldn't it be nice to avoid this obvious duplication of effort? Our main theorem for this section follows, and it mimics this previous example, while also avoiding all the overhead.

Theorem CINM

Computing the Inverse of a Nonsingular Matrix

Suppose A is a nonsingular square matrix of size n. Create the $n \times 2n$ matrix M by placing the $n \times n$ identity matrix I_n to the right of the matrix A. Let N be a matrix that is row-equivalent to M and in reduced row-echelon form. Finally, let J be the matrix formed from the final n columns of N. Then $AJ = I_n$.

Proof A is nonsingular, so by Theorem NMRRI [176] there is a sequence of row operations that will convert A into I_n . It is this same sequence of row operations that will convert M into N, since having the identity matrix in the first n columns of N is sufficient to guarantee that N is in reduced row-echelon form.

If we consider the systems of linear equations, $\mathcal{LS}(A, \mathbf{e}_i)$, $1 \leq i \leq n$, we see that the aforementioned sequence of row operations will also bring the augmented matrix of each of these systems into reduced row-echelon form. Furthermore, the unique solution to $\mathcal{LS}(A, \mathbf{e}_i)$ appears in column n+1 of the row-reduced augmented matrix of the system and is identical to column n+i of N. Let $\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \ldots, \mathbf{N}_{2n}$ denote the columns of N. So we find,

$$AJ = A[\mathbf{N}_{n+1}|\mathbf{N}_{n+2}|\mathbf{N}_{n+3}|\dots|\mathbf{N}_{n+n}]$$

$$= [A\mathbf{N}_{n+1}|A\mathbf{N}_{n+2}|A\mathbf{N}_{n+3}|\dots|A\mathbf{N}_{n+n}]$$
 Definition MM [468]
$$= [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\dots|\mathbf{e}_n]$$

$$= I_n$$
 Definition IM [175]

as desired.

We have to be just a bit careful here about both what this theorem says and what it doesn't say. If A is a nonsingular matrix, then we are guaranteed a matrix B such that $AB = I_n$, and the proof gives us a process for constructing B. However, the definition of the inverse of a matrix (Definition MI [506]) requires that $BA = I_n$ also. So at this juncture we must compute the matrix product in the "opposite" order before we claim B as the inverse of A. However, we'll soon see that this is *always* the case, in Theorem OSIS [541], so the title of this theorem is not inaccurate.

What if A is singular? At this point we only know that Theorem CINM [516] cannot be applied. The question of A's inverse is still open. (But see Theorem NI [542] in the next section.) We'll finish by computing the inverse for the coefficient matrix of Archetype B [1652], the one we just pulled from a hat in Example SABMI [503]. There are more examples in the Archetypes (Appendix A [1637]) to practice with, though notice that it is silly to ask for the inverse of a rectangular matrix (the sizes aren't right) and not every square matrix has an inverse (remember Example MWIAA [506]?).

Example CMIAB Computing a matrix inverse, Archetype B Archetype B [1652] has a coefficient matrix given as

$$B = \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix}$$

Exercising Theorem CINM [516] we set

$$M = \begin{bmatrix} -7 & -6 & -12 & 1 & 0 & 0\\ 5 & 5 & 7 & 0 & 1 & 0\\ 1 & 0 & 4 & 0 & 0 & 1 \end{bmatrix}.$$

which row reduces to

$$N = \begin{bmatrix} 1 & 0 & 0 & -10 & -12 & -9 \\ 0 & 1 & 0 & \frac{13}{2} & 8 & \frac{11}{2} \\ 0 & 0 & 1 & \frac{5}{2} & 3 & \frac{5}{2} \end{bmatrix}.$$

 So

$$B^{-1} = \begin{bmatrix} -10 & -12 & -9\\ \frac{13}{2} & 8 & \frac{11}{2}\\ \frac{5}{2} & 3 & \frac{5}{2} \end{bmatrix}$$

once we check that $B^{-1}B = I_3$ (the product in the opposite order is a consequence of the theorem).

While we can use a row-reducing procedure to compute any needed inverse, most computational devices have a built-in procedure to compute the inverse of a matrix straightaway. See: Computation MI.MMA [1576] Computation MI.SAGE [1590]

Subsection PMI Properties of Matrix Inverses

The inverse of a matrix enjoys some nice properties. We collect a few here. First, a matrix can have but one inverse.

Theorem MIU Matrix Inverse is Unique

Suppose the square matrix A has an inverse. Then A^{-1} is unique.

Proof As described in Technique U [1625], we will assume that A has two inverses. The hypothesis tells there is at least one. Suppose then that B and C are both inverses for A, so we know by Definition MI [506] that $AB = BA = I_n$ and $AC = CA = I_n$. Then we have,

$B = BI_n$	Theorem MMIM $[475]$
= B(AC)	Definition MI [506]

 \square

= (BA)C	Theorem MMA [478]
$=I_nC$	Definition MI [506]
= C	Theorem MMIM [475]

So we conclude that B and C are the same, and cannot be different. So any matrix that acts like *an* inverse, must be *the* inverse.

When most of us dress in the morning, we put on our socks first, followed by our shoes. In the evening we must then first remove our shoes, followed by our socks. Try to connect the conclusion of the following theorem with this everyday example.

Theorem SS

Socks and Shoes

Suppose A and B are invertible matrices of size n. Then AB is an invertible matrix and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof At the risk of carrying our everyday analogies too far, the proof of this theorem is quite easy when we compare it to the workings of a dating service. We have a statement about the inverse of the matrix AB, which for all we know right now might not even exist. Suppose AB was to sign up for a dating service with two requirements for a compatible date. Upon multiplication on the left, and on the right, the result should be the identity matrix. In other words, AB's ideal date would be its inverse.

Now along comes the matrix $B^{-1}A^{-1}$ (which we know exists because our hypothesis says both A and B are invertible and we can form the product of these two matrices), also looking for a date. Let's see if $B^{-1}A^{-1}$ is a good match for AB. First they meet at a non-committal neutral location, say a coffee shop, for quiet conversation:

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B$$

$$= B^{-1}I_nB$$

$$= B^{-1}B$$

$$= I_n$$
Theorem MMIM [478]
Theorem MMIM [475]
Definition MI [506]

The first date having gone smoothly, a second, more serious, date is arranged, say dinner and a show:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1}$$

Theorem MMA [478]
$$= AI_nA^{-1}$$

Definition MI [506]
$$= AA^{-1}$$

Theorem MMIM [475]
$$= I_n$$

Definition MI [506]

So the matrix $B^{-1}A^{-1}$ has met all of the requirements to be AB's inverse (date) and with the ensuing marriage proposal we can announce that $(AB)^{-1} = B^{-1}A^{-1}$.

Theorem MIMI

Matrix Inverse of a Matrix Inverse

Suppose A is an invertible matrix. Then A^{-1} is invertible and $(A^{-1})^{-1} = A$.

Proof As with the proof of Theorem SS [520], we examine if A is a suitable inverse for

 A^{-1} (by definition, the opposite is true).

$$AA^{-1} = I_n$$
 Definition MI [506]

and

$$A^{-1}A = I_n$$
 Definition MI [506]

The matrix A has met all the requirements to be the inverse of A^{-1} , and so is invertible and we can write $A = (A^{-1})^{-1}$.

Theorem MIT Matrix Inverse of a Transpose

Suppose A is an invertible matrix. Then A^t is invertible and $(A^t)^{-1} = (A^{-1})^t$.

Proof As with the proof of Theorem SS [520], we see if $(A^{-1})^t$ is a suitable inverse for A^t . Apply Theorem MMT [482] to see that

$(A^{-1})^t A^t = (AA^{-1})^t$	Theorem MMT [482]
$=I_{n}^{t}$	Definition MI [506]
$=I_n$	Definition SYM [434]

and

$$A^{t}(A^{-1})^{t} = (A^{-1}A)^{t}$$

Theorem MMT [482]

$=I_n^t$	Definition MI [506]
$=I_n$	Definition SYM [434]

The matrix $(A^{-1})^t$ has met all the requirements to be the inverse of A^t , and so is invertible and we can write $(A^t)^{-1} = (A^{-1})^t$.

Theorem MISM Matrix Inverse of a Scalar Multiple

Suppose A is an invertible matrix and α is a nonzero scalar. Then $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$ and αA is invertible.

Proof As with the proof of Theorem SS [520], we see if $\frac{1}{\alpha}A^{-1}$ is a suitable inverse for αA .

$$\left(\frac{1}{\alpha}A^{-1}\right)(\alpha A) = \left(\frac{1}{\alpha}\alpha\right)(AA^{-1})$$
$$= 1I_n$$
$$= I_n$$

Theorem MMSMM [477]

Scalar multiplicative inverses Property OM [431]

and

$$(\alpha A) \left(\frac{1}{\alpha} A^{-1}\right) = \left(\alpha \frac{1}{\alpha}\right) \left(A^{-1} A\right)$$
$$= 1I_n$$
$$= I_n$$

Theorem MMSMM [477]

Scalar multiplicative inverses Property OM [431] The matrix $\frac{1}{\alpha}A^{-1}$ has met all the requirements to be the inverse of αA , so we can write $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$.

Notice that there are some likely theorems that are missing here. For example, it would be tempting to think that $(A + B)^{-1} = A^{-1} + B^{-1}$, but this is false. Can you find a counterexample? (See Exercise MISLE.T10 [530].)

Subsection READ Reading Questions

1. Compute the inverse of the matrix below.

$$\begin{bmatrix} -2 & 3 \\ -3 & 4 \end{bmatrix}$$

2. Compute the inverse of the matrix below.

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & -2 & -3 \\ -2 & 4 & 6 \end{bmatrix}$$

3. Explain why Theorem SS [520] has the title it does. (Do not just state the theorem, explain the choice of the title making reference to the theorem itself.)

Subsection EXC Exercises

C16 If it exists, find the inverse of $A = \begin{vmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & -1 & 1 \end{vmatrix}$, and check your answer. Contributed by Chris Black Solution [531] C17 If it exists, find the inverse of $A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 2 & 1 \\ 3 & 1 & 2 \end{bmatrix}$, and check your answer. Contributed by Chris Black Solution [531] **C18** If it exists, find the inverse of $A = \begin{bmatrix} 1 & 3 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 1 \end{bmatrix}$, and check your answer. Contributed by Chris Black Solution [531] **C19** If it exists, find the inverse of $A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & 2 & 1 \\ 2 & 2 & 1 \end{bmatrix}$, and check your answer. Contributed by Chris Black Solution 531

C21 Verify that B is the inverse of A.

$$A = \begin{bmatrix} 1 & 1 & -1 & 2 \\ -2 & -1 & 2 & -3 \\ 1 & 1 & 0 & 2 \\ -1 & 2 & 0 & 2 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 4 & 2 & 0 & -1 \\ 8 & 4 & -1 & -1 \\ -1 & 0 & 1 & 0 \\ -6 & -3 & 1 & 1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [532]

C22 Recycle the matrices A and B from Exercise MISLE.C21 [526] and set

$$\mathbf{c} = \begin{bmatrix} 2\\1\\-3\\2 \end{bmatrix} \qquad \qquad \mathbf{d} = \begin{bmatrix} 1\\1\\1\\1\\1 \end{bmatrix}$$

Employ the matrix B to solve the two linear systems $\mathcal{LS}(A, \mathbf{c})$ and $\mathcal{LS}(A, \mathbf{d})$. Contributed by Robert Beezer Solution [532]

C23 If it exists, find the inverse of the 2×2 matrix

$$A = \begin{bmatrix} 7 & 3\\ 5 & 2 \end{bmatrix}$$

and check your answer. (See Theorem TTMI [509].) Contributed by Robert Beezer C24 If it exists, find the inverse of the 2×2 matrix

$$A = \begin{bmatrix} 6 & 3\\ 4 & 2 \end{bmatrix}$$

and check your answer. (See Theorem TTMI [509].) Contributed by Robert Beezer

C25 At the conclusion of Example CMI [512], verify that $BA = I_5$ by computing the matrix product.

Contributed by Robert Beezer

C26 Let

$$D = \begin{bmatrix} 1 & -1 & 3 & -2 & 1 \\ -2 & 3 & -5 & 3 & 0 \\ 1 & -1 & 4 & -2 & 2 \\ -1 & 4 & -1 & 0 & 4 \\ 1 & 0 & 5 & -2 & 5 \end{bmatrix}$$

Compute the inverse of D, D^{-1} , by forming the 5 × 10 matrix $[D | I_5]$ and row-reducing (Theorem CINM [516]). Then use a calculator to compute D^{-1} directly. Contributed by Robert Beezer Solution [532] C27 Let

$$E = \begin{bmatrix} 1 & -1 & 3 & -2 & 1 \\ -2 & 3 & -5 & 3 & -1 \\ 1 & -1 & 4 & -2 & 2 \\ -1 & 4 & -1 & 0 & 2 \\ 1 & 0 & 5 & -2 & 4 \end{bmatrix}$$

Compute the inverse of E, E^{-1} , by forming the 5 × 10 matrix $[E | I_5]$ and row-reducing (Theorem CINM [516]). Then use a calculator to compute E^{-1} directly. Contributed by Robert Beezer Solution [533]

C28 Let

$$C = \begin{bmatrix} 1 & 1 & 3 & 1 \\ -2 & -1 & -4 & -1 \\ 1 & 4 & 10 & 2 \\ -2 & 0 & -4 & 5 \end{bmatrix}$$

Compute the inverse of C, C^{-1} , by forming the 4×8 matrix $[C \mid I_4]$ and row-reducing (Theorem CINM [516]). Then use a calculator to compute C^{-1} directly. Contributed by Robert Beezer Solution [533]

C40 Find all solutions to the system of equations below, making use of the matrix inverse found in Exercise MISLE.C28 [528].

$$x_1 + x_2 + 3x_3 + x_4 = -4$$
$$-2x_1 - x_2 - 4x_3 - x_4 = 4$$

$$x_1 + 4x_2 + 10x_3 + 2x_4 = -20$$
$$-2x_1 - 4x_3 + 5x_4 = 9$$

Contributed by Robert Beezer Solution [533]

C41 Use the inverse of a matrix to find all the solutions to the following system of equations.

$$x_1 + 2x_2 - x_3 = -3$$

$$2x_1 + 5x_2 - x_3 = -4$$

$$-x_1 - 4x_2 = 2$$

Contributed by Robert Beezer Solution [534]

C42 Use a matrix inverse to solve the linear system of equations.

$$x_1 - x_2 + 2x_3 = 5$$

$$x_1 - 2x_3 = -8$$

$$2x_1 - x_2 - x_3 = -6$$

Contributed by Robert Beezer Solution [535]

T10 Construct an example to demonstrate that $(A + B)^{-1} = A^{-1} + B^{-1}$ is not true for all square matrices A and B of the same size.

Contributed by Robert Beezer Solution [536]

Subsection SOL Solutions

C16 Contributed by Chris Black Statement [525]
Answer:
$$A^{-1} = \begin{bmatrix} -2 & 1 & 1 \\ -1 & 1 & 0 \\ 3 & -1 & -1 \end{bmatrix}$$
.

C17 Contributed by Chris Black Statement [525]

The procedure we have for finding a matrix inverse fails for this matrix A since A does not row-reduce to I_3 . We suspect in this case that A is not invertible, although we do not yet know that concretely. (Stay tuned for upcoming revelations in Section MINM [537]!)

C18 Contributed by Chris Black Statement [525]
Answer:
$$A^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & -1 & 0 \\ -2 & 4 & -1 \end{bmatrix}$$

C19 Contributed by Chris Black Statement [525]
Answer: $A^{-1} = \begin{bmatrix} 0 & -1/2 & 1/2 \\ 1 & -1/2 & -1/2 \\ -2 & 2 & 1 \end{bmatrix}$

C21 Contributed by Robert Beezer Statement [526]

Check that *both* matrix products (Definition MM [468]) AB and BA equal the 4×4 identity matrix I_4 (Definition IM [175]).

C22 Contributed by Robert Beezer Statement [526] Represent each of the two systems by a vector equality, $A\mathbf{x} = \mathbf{c}$ and $A\mathbf{y} = \mathbf{d}$. Then in the spirit of Example SABMI [503], solutions are given by

$$\mathbf{x} = B\mathbf{c} = \begin{bmatrix} 8\\21\\-5\\-16 \end{bmatrix} \qquad \qquad \mathbf{y} = B\mathbf{d} = \begin{bmatrix} 5\\10\\0\\-7 \end{bmatrix}$$

Notice how we could solve many more systems having A as the coefficient matrix, and how each such system has a unique solution. You might check your work by substituting the solutions back into the systems of equations, or forming the linear combinations of the columns of A suggested by Theorem SLSLC [229].

C26 Contributed by Robert Beezer Statement [527] The inverse of D is

$$D^{-1} = \begin{bmatrix} -7 & -6 & -3 & 2 & 1 \\ -7 & -4 & 2 & 2 & -1 \\ -5 & -2 & 3 & 1 & -1 \\ -6 & -3 & 1 & 1 & 0 \\ 4 & 2 & -2 & -1 & 1 \end{bmatrix}$$

C27 Contributed by Robert Beezer Statement [528]

The matrix E has no inverse, though we do not yet have a theorem that allows us to reach this conclusion. However, when row-reducing the matrix $[E \mid I_5]$, the first 5 columns will not row-reduce to the 5 × 5 identity matrix, so we are a t a loss on how we might compute the inverse. When requesting that your calculator compute E^{-1} , it should give some indication that E does not have an inverse.

C28 Contributed by Robert Beezer Statement [528] Employ Theorem CINM [516],

Γ	1	1	3	1	1	0	0	0]		$\left[1 \right]$	0	0	0	38	18	-5	-2
	-2	-1	-4	-1	0	1	0	0	RREF	0	1	0	0	96	47	-12	-5
	1	4	10	2	0	0	1	0	\rightarrow	0	0	1	0	-39	-19	5	2
	-2	0	-4	5	0	0	0	1		0	0	0	1	-16	-8	2	1

And therefore we see that C is nonsingular (C row-reduces to the identity matrix, Theorem NMRRI [176]) and by Theorem CINM [516],

$$C^{-1} = \begin{bmatrix} 38 & 18 & -5 & -2\\ 96 & 47 & -12 & -5\\ -39 & -19 & 5 & 2\\ -16 & -8 & 2 & 1 \end{bmatrix}$$

C40 Contributed by Robert Beezer Statement [528] View this system as $\mathcal{LS}(C, \mathbf{b})$, where C is the 4 × 4 matrix from Exercise MISLE.C28 [528] and $\mathbf{b} = \begin{bmatrix} -4 \\ 4 \\ -20 \\ 9 \end{bmatrix}$. Since *C* was seen to be nonsingular in Exercise MISLE.C28 [528] Theorem

SNCM [543] says the solution, which is unique by Theorem NMUS [180], is given by

$$C^{-1}\mathbf{b} = \begin{bmatrix} 38 & 18 & -5 & -2\\ 96 & 47 & -12 & -5\\ -39 & -19 & 5 & 2\\ -16 & -8 & 2 & 1 \end{bmatrix} \begin{bmatrix} -4\\ 4\\ -20\\ 9 \end{bmatrix} = \begin{bmatrix} 2\\ -1\\ -2\\ 1 \end{bmatrix}$$

Notice that this solution can be easily checked in the original system of equations.

C41 Contributed by Robert Beezer Statement [529] The coefficient matrix of this system of equations is

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \\ -1 & -4 & 0 \end{bmatrix}$$

and the vector of constants is $\mathbf{b} = \begin{bmatrix} -3 \\ -4 \\ 2 \end{bmatrix}$. So by Theorem SLEMM [462] we can convert

the system to the form $A\mathbf{x} = \mathbf{b}$. Row-reducing this matrix yields the identity matrix so by Theorem NMRRI [176] we know A is nonsingular. This allows us to apply Theorem SNCM [543] to find the unique solution as

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -4 & 4 & 3\\ 1 & -1 & -1\\ -3 & 2 & 1 \end{bmatrix} \begin{bmatrix} -3\\ -4\\ 2 \end{bmatrix} = \begin{bmatrix} 2\\ -1\\ 3 \end{bmatrix}$$

Remember, you can check this solution easily by evaluating the matrix-vector product $A\mathbf{x}$ (Definition MVP [461]).

C42 Contributed by Robert Beezer Statement [529] We can reformulate the linear system as a vector equality with a matrix-vector product via Theorem SLEMM [462]. The system is then represented by $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 0 & -2 \\ 2 & -1 & -1 \end{bmatrix} \qquad \qquad \mathbf{b} = \begin{bmatrix} 5 \\ -8 \\ -6 \end{bmatrix}$$

According to Theorem SNCM [543], if A is nonsingular then the (unique) solution will be given by $A^{-1}\mathbf{b}$. We attempt the computation of A^{-1} through Theorem CINM [516], or with our favorite computational device and obtain,

$$A^{-1} = \begin{bmatrix} 2 & 3 & -2 \\ 3 & 5 & -4 \\ 1 & 1 & -1 \end{bmatrix}$$

So by Theorem NI [542], we know A is nonsingular, and so the unique solution is

$$A^{-1}\mathbf{b} = \begin{bmatrix} 2 & 3 & -2 \\ 3 & 5 & -4 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ -8 \\ -6 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix}$$

T10 Contributed by Robert Beezer Statement [530]

For a large collection of small examples, let D be any 2×2 matrix that has an inverse (Theorem TTMI [509] can help you construct such a matrix, I_2 is a simple choice). Set A = D and B = (-1)D. While A^{-1} and B^{-1} both exist, what is $(A + B)^{-1}$?

For a large collection of examples of any size, consider $A = B = I_n$. Can the proposed statement be salvaged to become a theorem?

Section MINM Matrix Inverses and Nonsingular Matrices

We saw in Theorem CINM [516] that if a square matrix A is nonsingular, then there is a matrix B so that $AB = I_n$. In other words, B is halfway to being an inverse of A. We will see in this section that B automatically fulfills the second condition ($BA = I_n$). Example MWIAA [506] showed us that the coefficient matrix from Archetype A [1642] had no inverse. Not coincidentally, this coefficient matrix is singular. We'll make all these connections precise now. Not many examples or definitions in this section, just theorems.

Subsection NMI Nonsingular Matrices are Invertible

We need a couple of technical results for starters. Some books would call these minor, but essential, results "lemmas." We'll just call 'em theorems. See Technique LC [1634] for more on the distinction.

The first of these technical results is interesting in that the hypothesis says something about a product of two square matrices and the conclusion then says the same thing about each individual matrix in the product. This result has an analogy in the algebra of complex numbers: suppose α , $\beta \in \mathbb{C}$, then $\alpha\beta \neq 0$ if and only if $\alpha \neq 0$ and $\beta \neq 0$. We can view this result as suggesting that the term "nonsingular" for matrices is like the term "nonzero" for scalars.

Theorem NPNT

Nonsingular Product has Nonsingular Terms

Suppose that A and B are square matrices of size n. The product AB is nonsingular if and only if A and B are both nonsingular.

Proof (\Rightarrow) We'll do this portion of the proof in two parts, each as a proof by contradiction (Technique CD [1623]). Assume that AB is nonsingular. Establishing that B is nonsingular is the easier part, so we will do it first, but in reality, we will *need* to know that B is nonsingular when we prove that A is nonsingular.

You can also think of this proof as being a study of four possible conclusions in the table below. One of the four rows *must* happen (the list is exhaustive). In the proof we learn that the first three rows lead to contradictions, and so are impossible. That leaves the fourth row as a certainty, which is our desired conclusion.

A	В	Case
Singular	Singular	1
Nonsingular	Singular	1
Singular	Nonsingular	2
Nonsingular	Nonsingular	

Part 1. Suppose B is singular. Then there is a nonzero vector \mathbf{z} that is a solution to $\mathcal{LS}(B, \mathbf{0})$. So

$(AB)\mathbf{z} = A(B\mathbf{z})$	Theorem MMA $[478]$
= A 0	Theorem SLEMM [462]
= 0	Theorem MMZM $[474]$

Because \mathbf{z} is a nonzero solution to $\mathcal{LS}(AB, \mathbf{0})$, we conclude that AB is singular (Definition NM [173]). This is a contradiction, so B is nonsingular, as desired.

Part 2. Suppose A is singular. Then there is a nonzero vector \mathbf{y} that is a solution to $\mathcal{LS}(A, \mathbf{0})$. Now consider the linear system $\mathcal{LS}(B, \mathbf{y})$. Since we know B is nonsingular from Case 1, the system has a unique solution (Theorem NMUS [180]), which we will denote as \mathbf{w} . We first claim \mathbf{w} is not the zero vector either. Assuming the opposite, suppose that $\mathbf{w} = \mathbf{0}$ (Technique CD [1623]). Then

$\mathbf{y} = B\mathbf{w}$	Theorem SLEMM [462]
= B 0	Hypothesis
= 0	Theorem MMZM [474]

contrary to **y** being nonzero. So $\mathbf{w} \neq \mathbf{0}$. The pieces are in place, so here we go,

 $(AB)\mathbf{w} = A(B\mathbf{w})$ Theorem MMA [478]

= Ay Theorem SLEMM [462] = 0 Theorem SLEMM [462]

So **w** is a nonzero solution to $\mathcal{LS}(AB, \mathbf{0})$, and thus we can say that AB is singular (Definition NM [173]). This is a contradiction, so A is nonsingular, as desired.

(\Leftarrow) Now assume that both A and B are nonsingular. Suppose that $\mathbf{x} \in \mathbb{C}^n$ is a solution to $\mathcal{LS}(AB, \mathbf{0})$. Then

$0 = (AB)\mathbf{x}$	Theorem SLEMM [462]
$=A\left(B\mathbf{x} ight)$	Theorem MMA [478]

By Theorem SLEMM [462], $B\mathbf{x}$ is a solution to $\mathcal{LS}(A, \mathbf{0})$, and by the definition of a nonsingular matrix (Definition NM [173]), we conclude that $B\mathbf{x} = \mathbf{0}$. Now, by an entirely similar argument, the nonsingularity of B forces us to conclude that $\mathbf{x} = \mathbf{0}$. So the only solution to $\mathcal{LS}(AB, \mathbf{0})$ is the zero vector and we conclude that AB is nonsingular by Definition NM [173].

This is a powerful result in the "forward" direction, because it allows us to begin with a hypothesis that something complicated (the matrix product AB) has the property of being nonsingular, and we can then conclude that the simpler constituents (A and B individually) then also have the property of being nonsingular. If we had thought that the matrix product was an artificial construction, results like this would make us begin to think twice.

The contrapositive of this result is equally interesting. It says that A or B (or both) is a singular matrix if and only if the product AB is singular. Notice how the negation of the theorem's conclusion (A and B both nonsingular) becomes the statement "at least one of Aand B is singular." (See Technique CP [1621].)

Theorem OSIS One-Sided Inverse is Sufficient

Suppose A and B are square matrices of size n such that $AB = I_n$. Then $BA = I_n$.

Proof The matrix I_n is nonsingular (since it row-reduces easily to I_n , Theorem NMRRI [176]). So A and B are nonsingular by Theorem NPNT [538], so in particular B is non-singular. We can therefore apply Theorem CINM [516] to assert the existence of a matrix C so that $BC = I_n$. This application of Theorem CINM [516] could be a bit confusing, mostly because of the names of the matrices involved. B is nonsingular, so there must be a "right-inverse" for B, and we're calling it C.

Now

$$BA = (BA)I_n$$

$$= (BA)(BC)$$

$$= B(AB)C$$

$$= BI_nC$$

$$= BC$$

$$= I_n$$
Theorem MMIM [475]
Theorem MMA [478]
Hypothesis
Theorem MMIM [475]
Theorem MMIM [475]
Theorem CINM [516]

which is the desired conclusion.

So Theorem OSIS [541] tells us that if A is nonsingular, then the matrix B guaranteed by Theorem CINM [516] will be both a "right-inverse" and a "left-inverse" for A, so A is invertible and $A^{-1} = B$.

So if you have a nonsingular matrix, A, you can use the procedure described in Theorem CINM [516] to find an inverse for A. If A is singular, then the procedure in Theorem CINM [516] will fail as the first n columns of M will not row-reduce to the identity matrix. However, we can say a bit more. When A is singular, then A does not have an inverse (which is very different from saying that the procedure in Theorem CINM [516] fails to find an inverse). This may feel like we are splitting hairs, but it's important that we do not make unfounded assumptions. These observations motivate the next theorem.

Theorem NI

Nonsingularity is Invertibility

Suppose that A is a square matrix. Then A is nonsingular if and only if A is invertible. \Box

Proof (\Leftarrow) Since A is invertible, we can write $I_n = AA^{-1}$ (Definition MI [506]). Notice that I_n is nonsingular (Theorem NMRRI [176]) so Theorem NPNT [538] implies that A (and A^{-1}) is nonsingular.

(⇒) Suppose now that A is nonsingular. By Theorem CINM [516] we find B so that $AB = I_n$. Then Theorem OSIS [541] tells us that $BA = I_n$. So B is A's inverse, and by construction, A is invertible.

So for a square matrix, the properties of having an inverse and of having a trivial null space are one and the same. Can't have one without the other.

Theorem NME3 Nonsingular Matrix Equivalences, Round 3

Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.

Proof We can update our list of equivalences for nonsingular matrices (Theorem NME2 [330]) with the equivalent condition from Theorem NI [542]. ■

In the case that A is a nonsingular coefficient matrix of a system of equations, the inverse allows us to very quickly compute the unique solution, for any vector of constants.

Theorem SNCM Solution with Nonsingular Coefficient Matrix Suppose that A is nonsingular. Then the unique solution to $\mathcal{LS}(A, \mathbf{b})$ is $A^{-1}\mathbf{b}$.

Proof By Theorem NMUS [180] we know already that $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every choice of **b**. We need to show that the expression stated is indeed a solution (*the* solution). That's easy, just "plug it in" to the corresponding vector equation representation (Theorem SLEMM [462]),

$A\left(A^{-1}\mathbf{b}\right) = \left(AA^{-1}\right)\mathbf{b}$	Theorem MMA $[478]$
$=I_n\mathbf{b}$	Definition MI [506]
= b	Theorem MMIM [475]

Since $A\mathbf{x} = \mathbf{b}$ is true when we substitute $A^{-1}\mathbf{b}$ for \mathbf{x} , $A^{-1}\mathbf{b}$ is a (the!) solution to $\mathcal{LS}(A, \mathbf{b})$.

Subsection UM Unitary Matrices

Recall that the adjoint of a matrix is $A^* = (\overline{A})^t$ (Definition A [442]).

Definition UM Unitary Matrices Suppose that U is a square matrix of size n such that $U^*U = I_n$. Then we say U is **unitary**.

 \triangle

This condition may seem rather far-fetched at first glance. Would there be *any* matrix that behaved this way? Well, yes, here's one.

Example UM3 Unitary matrix of size 3

$$U = \begin{bmatrix} \frac{1+i}{\sqrt{5}} & \frac{3+2i}{\sqrt{55}} & \frac{2+2i}{\sqrt{22}} \\ \frac{1-i}{\sqrt{5}} & \frac{2+2i}{\sqrt{55}} & \frac{-3+i}{\sqrt{22}} \\ \frac{i}{\sqrt{5}} & \frac{3-5i}{\sqrt{55}} & -\frac{2}{\sqrt{22}} \end{bmatrix}$$

The computations get a bit tiresome, but if you work your way through the computation of U^*U , you *will* arrive at the 3×3 identity matrix I_3 .

Unitary matrices do not have to look quite so gruesome. Here's a larger one that is a bit more pleasing.

Example UPM Unitary permutation matrix The matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

is unitary as can be easily checked. Notice that it is just a rearrangement of the columns of the 5×5 identity matrix, I_5 (Definition IM [175]).

An interesting exercise is to build another 5×5 unitary matrix, R, using a different rearrangement of the columns of I_5 . Then form the product PR. This will be another unitary matrix (Exercise MINM.T10 [554]). If you were to build all $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$ matrices of this type you would have a set that remains closed under matrix multiplication. It is an example of another algebraic structure known as a **group** since together the set and the one operation (matrix multiplication here) is closed, associative, has an identity (I_5), and inverses (Theorem UMI [546]). Notice though that the operation in this group is not commutative!

If a matrix A has only real number entries (we say it is a **real matrix**) then the defining property of being unitary simplifies to $A^t A = I_n$. In this case we, and everybody else, call the matrix **orthogonal**, so you may often encounter this term in your other reading when the complex numbers are not under consideration.

Unitary matrices have easily computed inverses. They also have columns that form orthonormal sets. Here are the theorems that show us that unitary matrices are not as strange as they might initially appear.

Theorem UMI

Unitary Matrices are Invertible

Suppose that U is a unitary matrix of size n. Then U is nonsingular, and $U^{-1} = U^*$.

Proof By Definition UM [544], we know that $U^*U = I_n$. The matrix I_n is nonsingular (since it row-reduces easily to I_n , Theorem NMRRI [176]). So by Theorem NPNT [538], U

 \square

and U^* are both nonsingular matrices.

The equation $U^*U = I_n$ gets us halfway to an inverse of U, and Theorem OSIS [541] tells us that then $UU^* = I_n$ also. So U and U^* are inverses of each other (Definition MI [506]).

Theorem CUMOS

Columns of Unitary Matrices are Orthonormal Sets

m

Suppose that A is a square matrix of size n with columns $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \dots, \mathbf{A}_n\}$. Then A is a unitary matrix if and only if S is an orthonormal set.

Proof The proof revolves around recognizing that a typical entry of the product A^*A is an inner product of columns of A. Here are the details to support this claim.

$$[A^*A]_{ij} = \sum_{k=1}^{n} [A^*]_{ik} [A]_{kj} \qquad \text{Theorem EMP [471]}$$
$$= \sum_{k=1}^{n} \left[\overline{A}^t\right]_{ik} [A]_{kj} \qquad \text{Theorem EMP [471]}$$
$$= \sum_{k=1}^{n} \left[\overline{A}\right]_{ki} [A]_{kj} \qquad \text{Definition TM [433]}$$
$$= \sum_{k=1}^{n} \overline{[A]_{ki}} [A]_{kj} \qquad \text{Definition CCM [438]}$$

$$= \sum_{k=1}^{n} \overline{[\mathbf{A}_{i}]_{k}} [\mathbf{A}_{j}]_{k}$$
$$= \langle \mathbf{A}_{i}, \mathbf{A}_{j} \rangle \qquad \qquad \text{Definition IP [398]}$$

We now employ this equality in a chain of equivalences,

$$S = \{\mathbf{A}_{1}, \mathbf{A}_{2}, \mathbf{A}_{3}, \dots, \mathbf{A}_{n}\} \text{ is an orthonormal set}$$

$$\iff \langle \mathbf{A}_{i}, \mathbf{A}_{j} \rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$$\iff [A^{*}A]_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

$$\iff [A^{*}A]_{ij} = [I_{n}]_{ij}, \ 1 \leq i \leq n, \ 1 \leq j \leq n \end{cases}$$

$$\implies A^{*}A = I_{n}$$

$$\iff A \text{ is a unitary matrix}$$
Definition IM [175]
Definition ME [427]
Definition UM [544]

Example OSMC Orthonormal set from matrix columns The matrix

$$U = \begin{bmatrix} \frac{1+i}{\sqrt{5}} & \frac{3+2i}{\sqrt{55}} & \frac{2+2i}{\sqrt{22}} \\ \frac{1-i}{\sqrt{5}} & \frac{2+2i}{\sqrt{55}} & \frac{-3+i}{\sqrt{22}} \\ \frac{i}{\sqrt{5}} & \frac{3-5i}{\sqrt{55}} & -\frac{2}{\sqrt{22}} \end{bmatrix}$$

from Example UM3 [545] is a unitary matrix. By Theorem CUMOS [547], its columns

$$\left\{ \begin{bmatrix} \frac{1+i}{\sqrt{5}} \\ \frac{1-i}{\sqrt{5}} \\ \frac{i}{\sqrt{5}} \end{bmatrix}, \begin{bmatrix} \frac{3+2i}{\sqrt{55}} \\ \frac{2+2i}{\sqrt{55}} \\ \frac{3-5i}{\sqrt{55}} \end{bmatrix}, \begin{bmatrix} \frac{2+2i}{\sqrt{22}} \\ -\frac{3+i}{\sqrt{22}} \\ -\frac{2}{\sqrt{22}} \end{bmatrix} \right\}$$

form an orthonormal set. You might find checking the six inner products of pairs of these vectors easier than doing the matrix product U^*U . Or, because the inner product is anticommutative (Theorem IPAC [402]) you only need check three inner products (see Exercise MINM.T12 [555]).

When using vectors and matrices that only have real number entries, orthogonal matrices are those matrices with inverses that equal their transpose. Similarly, the inner product is the familiar dot product. Keep this special case in mind as you read the next theorem.

Theorem UMPIP

Unitary Matrices Preserve Inner Products

Suppose that U is a unitary matrix of size n and **u** and **v** are two vectors from \mathbb{C}^n . Then

 $\langle U\mathbf{u}, U\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$ and $\|U\mathbf{v}\| = \|\mathbf{v}\|$

\mathbf{Proof}

$$\langle U\mathbf{u}, U\mathbf{v} \rangle = \left(\overline{U}\mathbf{u}\right)^t U\mathbf{v}$$

$$= \left(\overline{U}\overline{\mathbf{u}}\right)^t U\mathbf{v}$$

$$= \overline{\mathbf{u}}^t \overline{U}^t U\mathbf{v}$$

$$= \overline{\mathbf{u}}^t U^* U\mathbf{v}$$

$$= \overline{\mathbf{u}}^t I_n \mathbf{v}$$

$$= \overline{\mathbf{u}}^t \mathbf{v}$$

$$= \langle \mathbf{u}, \mathbf{v} \rangle$$

Theorem MMIP [480] Theorem MMCC [481] Theorem MMT [482] Definition A [442] Definition UM [544] Theorem MMIM [475] Theorem MMIP [480]

The second conclusion is just a specialization of the first conclusion.

$$\|U\mathbf{v}\| = \sqrt{\|U\mathbf{v}\|^2}$$

= $\sqrt{\langle U\mathbf{v}, U\mathbf{v} \rangle}$ Theorem IPN [406]
= $\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$
= $\sqrt{\|\mathbf{v}\|^2}$ Theorem IPN [406]
= $\|\mathbf{v}\|$

 \square

Aside from the inherent interest in this theorem, it makes a bigger statement about unitary matrices. When we view vectors geometrically as directions or forces, then the norm equates to a notion of length. If we transform a vector by multiplication with a unitary matrix, then the length (norm) of that vector stays the same. If we consider column vectors with two or three slots containing only real numbers, then the inner product of two such vectors is just the dot product, and this quantity can be used to compute the angle between two vectors. When two vectors are multiplied (transformed) by the same unitary matrix, their dot product is unchanged and their individual lengths are unchanged. This results in the angle between the two vectors remaining unchanged.

A "unitary transformation" (matrix-vector products with unitary matrices) thus preserve geometrical relationships among vectors representing directions, forces, or other physical quantities. In the case of a two-slot vector with real entries, this is simply a rotation. These sorts of computations are exceedingly important in computer graphics such as games and real-time simulations, especially when increased realism is achieved by performing many such computations quickly. We will see unitary matrices again in subsequent sections (especially Theorem OD [1429]) and in each instance, consider the interpretation of the unitary matrix as a sort of geometry-preserving transformation. Some authors use the term **isometry** to highlight this behavior. We will speak loosely of a unitary matrix as being a sort of generalized rotation.

A final reminder: the terms "dot product," "symmetric matrix" and "orthogonal matrix" used in reference to vectors or matrices with real number entries correspond to the terms

"inner product," "Hermitian matrix" and "unitary matrix" when we generalize to include complex number entries, so keep that in mind as you read elsewhere.

Subsection READ Reading Questions

1. Compute the inverse of the coefficient matrix of the system of equations below and use the inverse to solve the system.

$$4x_1 + 10x_2 = 12 2x_1 + 6x_2 = 4$$

2. In the reading questions for Section MISLE [503] you were asked to find the inverse of the 3×3 matrix below.

$$\begin{bmatrix} 2 & 3 & 1 \\ 1 & -2 & -3 \\ -2 & 4 & 6 \end{bmatrix}$$

Because the matrix was not nonsingular, you had no theorems at that point that would allow you to compute the inverse. Explain why you now know that the inverse does not exist (which is different than not being able to compute it) by quoting the relevant theorem's acronym. 3. Is the matrix A unitary? Why?

$$A = \begin{bmatrix} \frac{1}{\sqrt{22}} \left(4+2i\right) & \frac{1}{\sqrt{374}} \left(5+3i\right) \\ \frac{1}{\sqrt{22}} \left(-1-i\right) & \frac{1}{\sqrt{374}} \left(12+14i\right) \end{bmatrix}$$

Subsection EXC Exercises

C20 Let
$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}$$
 and $B = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. Verify that AB is nonsingular. Contributed by Chris Black

C40 Solve the system of equations below using the inverse of a matrix.

$$x_1 + x_2 + 3x_3 + x_4 = 5$$

-2x₁ - x₂ - 4x₃ - x₄ = -7
x₁ + 4x₂ + 10x₃ + 2x₄ = 9
-2x₁ - 4x₃ + 5x₄ = 9

Contributed by Robert Beezer Solution [556]

M10 Find values of
$$x, y, z$$
 so that matrix $A = \begin{bmatrix} 1 & 2 & x \\ 3 & 0 & y \\ 1 & 1 & z \end{bmatrix}$ is invertible.
Contributed by Chris Black Solution [557]

M11 Find values of
$$x, y z$$
 so that matrix $A = \begin{bmatrix} 1 & x & 1 \\ 1 & y & 4 \\ 0 & z & 5 \end{bmatrix}$ is singular.
Contributed by Chris Black Solution [558]

M15 If A and B are $n \times n$ matrices, A is nonsingular, and B is singular, show directly that AB is singular, without using Theorem NPNT [538]. Contributed by Chris Black Solution [558]

M20 Construct an example of a 4×4 unitary matrix. Contributed by Robert Beezer Solution [556]

M80 Matrix multiplication interacts nicely with many operations. But not always with transforming a matrix to reduced row-echelon form. Suppose that A is an $m \times n$ matrix and B is an $n \times p$ matrix. Let P be a matrix that is row-equivalent to A and in reduced row-echelon form, Q be a matrix that is row-equivalent to B and in reduced row-echelon form, and let R be a matrix that is row-equivalent to AB and in reduced row-echelon form. Is PQ = R? (In other words, with nonstandard notation, is rref(A)rref(B) = rref(AB)?)

Construct a counterexample to show that, in general, this statement is false. Then find a large class of matrices where if A and B are in the class, then the statement is true. Contributed by Mark Hamrick Solution [558]

T10 Suppose that Q and P are unitary matrices of size n. Prove that QP is a unitary matrix.

Contributed by Robert Beezer

T11 Prove that Hermitian matrices (Definition HM [486]) have real entries on the diagonal. More precisely, suppose that A is a Hermitian matrix of size n. Then $[A]_{ii} \in \mathbb{R}$, $1 \le i \le n$. Contributed by Robert Beezer

T12 Suppose that we are checking if a square matrix of size n is unitary. Show that a straightforward application of Theorem CUMOS [547] requires the computation of n^2 inner products when the matrix is unitary, and fewer when the matrix is not orthogonal. Then show that this maximum number of inner products can be reduced to $\frac{1}{2}n(n+1)$ in light of Theorem IPAC [402].

Contributed by Robert Beezer

T25 The notation A^k means a repeated matrix product between k copies of the square matrix A.

(a) Assume A is an $n \times n$ matrix where $A^2 = \mathcal{O}$ (which does not imply that $A = \mathcal{O}$.) Prove that $I_n - A$ is invertible by showing that $I_n + A$ is an inverse of $I_n - A$.

(b) Assume that A is an $n \times n$ matrix where $A^3 = \mathcal{O}$. Prove that $I_n - A$ is invertible.

(c) Form a general theorem based on your observations from parts (a) and (b) and provide a proof.

Contributed by Manley Perkel

Subsection SOL Solutions

C40 Contributed by Robert Beezer Statement [553] The coefficient matrix and vector of constants for the system are

$$\begin{bmatrix} 1 & 1 & 3 & 1 \\ -2 & -1 & -4 & -1 \\ 1 & 4 & 10 & 2 \\ -2 & 0 & -4 & 5 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 5 \\ -7 \\ 9 \\ 9 \end{bmatrix}$$

 A^{-1} can be computed by using a calculator, or by the method of Theorem CINM [516]. Then Theorem SNCM [543] says the unique solution is

$$A^{-1}\mathbf{b} = \begin{bmatrix} 38 & 18 & -5 & -2\\ 96 & 47 & -12 & -5\\ -39 & -19 & 5 & 2\\ -16 & -8 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5\\ -7\\ 9\\ 9 \end{bmatrix} = \begin{bmatrix} 1\\ -2\\ 1\\ 3 \end{bmatrix}$$

M20 Contributed by Robert Beezer Statement [554] The 4×4 identity matrix, I_4 , would be one example (Definition IM [175]). Any of the 23 other rearrangements of the columns of I_4 would be a simple, but less trivial, example. See Example UPM [545].

M10 Contributed by Chris Black Statement [553]

There are an infinite number of possible answers. We want to find a vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ so that the set

$$S = \left\{ \begin{bmatrix} 1\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\0\\1 \end{bmatrix}, \begin{bmatrix} x\\y\\z \end{bmatrix} \right\}$$

is a linearly independent set. We need a vector not in the span of the first two columns, which geometrically means that we need it to not be in the same plane as the first two columns of A. We can choose any values we want for x and y, and then choose a value of z that makes the three vectors independent.

I will (arbitrarily) choose x = 1, y = 1. Then, we have

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 1 \\ 1 & 1 & z \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2z - 1 \\ 0 & 1 & 1 - z \\ 0 & 0 & 4 - 6z \end{bmatrix}$$

which is invertible if and only if $4 - 6z \neq 0$. Thus, we can choose any value as long as $z \neq \frac{2}{3}$, so we choose z = 0, and we have found a matrix $A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ that is invertible.

M11 Contributed by Chris Black Statement [554] There are an infinite number of possible answers. We need the set of vectors

$$S = \left\{ \begin{bmatrix} 1\\1\\0 \end{bmatrix}, \begin{bmatrix} x\\y\\z \end{bmatrix}, \begin{bmatrix} 1\\4\\5 \end{bmatrix} \right\}$$

to be linearly dependent. One way to do this by inspection is to have $\begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} 1 \\ 4 \\ 5 \end{vmatrix}$. Thus, if

we let
$$x = 1, y = 4, z = 5$$
, then the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 4 & 4 \\ 0 & 5 & 5 \end{bmatrix}$ is singular.

M15 Contributed by Chris Black Statement [554] If B is singular, then there exists a vector $\mathbf{x} \neq \mathbf{0}$ so that $\mathbf{x} \in \mathcal{N}(B)$. Thus, $B\mathbf{x} = \mathbf{0}$, so $A(B\mathbf{x}) = (AB)\mathbf{x} = \mathbf{0}$, so $\mathbf{x} \in \mathcal{N}(AB)$. Since the null space of AB is not trivial, AB is a singular matrix.

M80 Contributed by Robert Beezer Statement [554] Take

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Then A is already in reduced row-echelon form, and by swapping rows, B row-reduces to A. So the product of the row-echelon forms of A is $AA = A \neq O$. However, the product AB is the 2×2 zero matrix, which is in reduced-echelon form, and not equal to AA. When you get there, Theorem PEEF [621] or Theorem EMDRO [886] might shed some light on why we would not expect this statement to be true in general.

If A and B are nonsingular, then AB is nonsingular (Theorem NPNT [538]), and all three matrices A, B and AB row-reduce to the identity matrix (Theorem NMRRI [176]). By Theorem MMIM [475], the desired relationship is true.

Section CRS Column and Row Spaces

Theorem SLSLC [229] showed us that there is a natural correspondence between solutions to linear systems and linear combinations of the columns of the coefficient matrix. This idea motivates the following important definition.

Definition CSM Column Space of a Matrix

Suppose that A is an $m \times n$ matrix with columns $\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n\}$. Then the **column space** of A, written $\mathcal{C}(A)$, is the subset of \mathbb{C}^m containing all linear combinations of the columns of A,

$$\mathcal{C}(A) = \langle \{\mathbf{A}_1, \, \mathbf{A}_2, \, \mathbf{A}_3, \, \dots, \, \mathbf{A}_n \} \rangle$$

(This definition contains Notation CSM.)

Some authors refer to the column space of a matrix as the **range**, but we will reserve this term for use with linear transformations (Definition RLT [1178]).

 \wedge

Subsection CSSE Column Spaces and Systems of Equations

Upon encountering any new set, the first question we ask is what objects are in the set, and which objects are not? Here's an example of one way to answer this question, and it will motivate a theorem that will then answer the question precisely.

Example CSMCS Column space of a matrix and consistent systems

Archetype D [1670] and Archetype E [1678] are linear systems of equations, with an identical 3×4 coefficient matrix, which we call A here. However, Archetype D [1670] is consistent, while Archetype E [1678] is not. We can explain this difference by employing the column space of the matrix A.

The column vector of constants, \mathbf{b} , in Archetype D [1670] is

$$\mathbf{b} = \begin{bmatrix} 8\\-12\\4 \end{bmatrix}$$

One solution to $\mathcal{LS}(A, \mathbf{b})$, as listed, is

$$\mathbf{x} = \begin{bmatrix} 7\\8\\1\\3 \end{bmatrix}$$

By Theorem SLSLC [229], we can summarize this solution as a linear combination of the columns of A that equals **b**,

$$7\begin{bmatrix}2\\-3\\1\end{bmatrix}+8\begin{bmatrix}1\\4\\1\end{bmatrix}+1\begin{bmatrix}7\\-5\\4\end{bmatrix}+3\begin{bmatrix}-7\\-6\\-5\end{bmatrix}=\begin{bmatrix}8\\-12\\4\end{bmatrix}=\mathbf{b}$$

This equation says that **b** is a linear combination of the columns of A, and then by Definition CSM [560], we can say that $\mathbf{b} \in \mathcal{C}(A)$.

On the other hand, Archetype E [1678] is the linear system $\mathcal{LS}(A, \mathbf{c})$, where the vector of constants is

$$\mathbf{c} = \begin{bmatrix} 2\\ 3\\ 2 \end{bmatrix}$$

and this system of equations is inconsistent. This means $\mathbf{c} \notin \mathcal{C}(A)$, for if it were, then it would equal a linear combination of the columns of A and Theorem SLSLC [229] would lead

us to a solution of the system $\mathcal{LS}(A, \mathbf{c})$.

So if we fix the coefficient matrix, and vary the vector of constants, we can sometimes find consistent systems, and sometimes inconsistent systems. The vectors of constants that lead to consistent systems are exactly the elements of the column space. This is the content of the next theorem, and since it is an equivalence, it provides an alternate view of the column space.

Theorem CSCS

Column Spaces and Consistent Systems

Suppose A is an $m \times n$ matrix and **b** is a vector of size m. Then $\mathbf{b} \in \mathcal{C}(A)$ if and only if $\mathcal{LS}(A, \mathbf{b})$ is consistent.

Proof (\Rightarrow) Suppose $\mathbf{b} \in \mathcal{C}(A)$. Then we can write \mathbf{b} as some linear combination of the columns of A. By Theorem SLSLC [229] we can use the scalars from this linear combination to form a solution to $\mathcal{LS}(A, \mathbf{b})$, so this system is consistent.

 (\Leftarrow) If $\mathcal{LS}(A, \mathbf{b})$ is consistent, there is a solution that may be used with Theorem SLSLC [229] to write **b** as a linear combination of the columns of A. This qualifies **b** for membership in $\mathcal{C}(A)$.

This theorem tells us that asking if the system $\mathcal{LS}(A, \mathbf{b})$ is consistent is exactly the same question as asking if \mathbf{b} is in the column space of A. Or equivalently, it tells us that the column space of the matrix A is precisely those vectors of constants, \mathbf{b} , that can be paired with A to create a system of linear equations $\mathcal{LS}(A, \mathbf{b})$ that is consistent.

Employing Theorem SLEMM [462] we can form the chain of equivalences

 $\mathbf{b} \in \mathcal{C}(A) \iff \mathcal{LS}(A, \mathbf{b})$ is consistent $\iff A\mathbf{x} = \mathbf{b}$ for some \mathbf{x}

Thus, an alternative (and popular) definition of the column space of an $m \times n$ matrix A is

$$\mathcal{C}(A) = \{ \mathbf{y} \in \mathbb{C}^m | \mathbf{y} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{C}^n \} = \{ A\mathbf{x} | \mathbf{x} \in \mathbb{C}^n \} \subseteq \mathbb{C}^m$$

We recognize this as saying create *all* the matrix vector products possible with the matrix A by letting **x** range over all of the possibilities. By Definition MVP [461] we see that this means take all possible linear combinations of the columns of A — precisely the definition of the column space (Definition CSM [560]) we have chosen.

Notice how this formulation of the column space looks very much like the definition of the null space of a matrix (Definition NSM [152]), but for a rectangular matrix the column vectors of $\mathcal{C}(A)$ and $\mathcal{N}(A)$ have different sizes, so the sets are very different.

Given a vector **b** and a matrix A it is now very mechanical to test if $\mathbf{b} \in \mathcal{C}(A)$. Form the linear system $\mathcal{LS}(A, \mathbf{b})$, row-reduce the augmented matrix, $[A \mid \mathbf{b}]$, and test for consistency with Theorem RCLS [120]. Here's an example of this procedure.

Example MCSM Membership in the column space of a matrix

Consider the column space of the 3×4 matrix A,

$$A = \begin{bmatrix} 3 & 2 & 1 & -4 \\ -1 & 1 & -2 & 3 \\ 2 & -4 & 6 & -8 \end{bmatrix}$$

We first show that $\mathbf{v} = \begin{bmatrix} 18 \\ -6 \\ 12 \end{bmatrix}$ is in the column space of $A, \mathbf{v} \in \mathcal{C}(A)$. Theorem CSCS [563] says we need only check the consistency of $\mathcal{LS}(A, \mathbf{v})$. Form the augmented matrix and

row-reduce,

$$\begin{bmatrix} 3 & 2 & 1 & -4 & 18 \\ -1 & 1 & -2 & 3 & -6 \\ 2 & -4 & 6 & -8 & 12 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & -2 & 6 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Without a leading 1 in the final column, Theorem RCLS [120] tells us the system is consistent and therefore by Theorem CSCS [563], $\mathbf{v} \in \mathcal{C}(A)$.

If we wished to demonstrate explicitly that \mathbf{v} is a linear combination of the columns of A, we can find a solution (any solution) of $\mathcal{LS}(A, \mathbf{v})$ and use Theorem SLSLC [229] to construct the desired linear combination. For example, set the free variables to $x_3 = 2$ and $x_4 = 1$. Then a solution has $x_2 = 1$ and $x_1 = 6$. Then by Theorem SLSLC [229],

$$\mathbf{v} = \begin{bmatrix} 18\\-6\\12 \end{bmatrix} = 6 \begin{bmatrix} 3\\-1\\2 \end{bmatrix} + 1 \begin{bmatrix} 2\\1\\-4 \end{bmatrix} + 2 \begin{bmatrix} 1\\-2\\6 \end{bmatrix} + 1 \begin{bmatrix} -4\\3\\-8 \end{bmatrix}$$

Now we show that $\mathbf{w} = \begin{bmatrix} 2\\1\\-3 \end{bmatrix}$ is not in the column space of $A, \mathbf{w} \notin \mathcal{C}(A)$. Theorem CSCS [563] says we need only check the consistency of $\mathcal{LS}(A, \mathbf{w})$. Form the augmented matrix

and row-reduce,

$$\begin{bmatrix} 3 & 2 & 1 & -4 & 2 \\ -1 & 1 & -2 & 3 & 1 \\ 2 & -4 & 6 & -8 & -3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & -2 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

With a leading 1 in the final column, Theorem RCLS [120] tells us the system is inconsistent and therefore by Theorem CSCS [563], $\mathbf{w} \notin \mathcal{C}(A)$.

Theorem CSCS [563] completes a collection of three theorems, and one definition, that deserve comment. Many questions about spans, linear independence, null space, column spaces and similar objects can be converted to questions about systems of equations (homogeneous or not), which we understand well from our previous results, especially those in Chapter SLE [2]. These previous results include theorems like Theorem RCLS [120] which allows us to quickly decide consistency of a system, and Theorem BNS [334] which allows us to describe solution sets for homogeneous systems compactly as the span of a linearly independent set of column vectors.

The table below lists these four definitions and theorems along with a brief reminder of the statement and an example of how the statement is used.

Definition NSM [152]	
Synopsis	Null space is solution set of homogeneous system
Example	General solution sets described by Theorem PSPHS [256]
Theorem SLSLC [229]	
Synopsis	Solutions for linear combinations with unknown scalars
Example	Deciding membership in spans
Theorem SLEMM [462]	
Synopsis	System of equations represented by matrix-vector product
Example	Solution to $\mathcal{LS}(A, \mathbf{b})$ is $A^{-1}\mathbf{b}$ when A is nonsingular
Theorem CSCS [563]	
Synopsis	Column space vectors create consistent systems
Example	Deciding membership in column spaces

Subsection CSSOC Column Space Spanned by Original Columns

So we have a foolproof, automated procedure for determining membership in $\mathcal{C}(A)$. While this works just fine a vector at a time, we would like to have a more useful description of the set $\mathcal{C}(A)$ as a whole. The next example will preview the first of two fundamental results about the column space of a matrix.

Example CSTW Column space, two ways

Consider the 5×7 matrix A,

$$\begin{bmatrix} 2 & 4 & 1 & -1 & 1 & 4 & 4 \\ 1 & 2 & 1 & 0 & 2 & 4 & 7 \\ 0 & 0 & 1 & 4 & 1 & 8 & 7 \\ 1 & 2 & -1 & 2 & 1 & 9 & 6 \\ -2 & -4 & 1 & 3 & -1 & -2 & -2 \end{bmatrix}$$

According to the definition (Definition CSM [560]), the column space of A is

$$\mathcal{C}(A) = \left\langle \left\{ \begin{bmatrix} 2\\1\\0\\1\\-2 \end{bmatrix}, \begin{bmatrix} 4\\2\\0\\2\\-4 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\-1\\1 \end{bmatrix}, \begin{bmatrix} -1\\0\\4\\2\\3 \end{bmatrix}, \begin{bmatrix} 1\\2\\1\\1\\-1 \end{bmatrix}, \begin{bmatrix} 4\\4\\8\\9\\-2 \end{bmatrix}, \begin{bmatrix} 4\\7\\7\\6\\-2 \end{bmatrix} \right\} \right\rangle$$

While this is a concise description of an infinite set, we might be able to describe the span with fewer than seven vectors. This is the substance of Theorem BS [376]. So we take these seven vectors and make them the columns of matrix, which is simply the original matrix A

again. Now we row-reduce,

$$\begin{bmatrix} 2 & 4 & 1 & -1 & 1 & 4 & 4 \\ 1 & 2 & 1 & 0 & 2 & 4 & 7 \\ 0 & 0 & 1 & 4 & 1 & 8 & 7 \\ 1 & 2 & -1 & 2 & 1 & 9 & 6 \\ -2 & -4 & 1 & 3 & -1 & -2 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 2 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivot columns are $D = \{1, 3, 4, 5\}$, so we can create the set

$$T = \left\{ \begin{bmatrix} 2\\1\\0\\1\\-2 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\-1\\1 \end{bmatrix}, \begin{bmatrix} -1\\0\\4\\2\\3 \end{bmatrix}, \begin{bmatrix} 1\\2\\1\\1\\-1 \end{bmatrix} \right\}$$

and know that $\mathcal{C}(A) = \langle T \rangle$ and T is a linearly independent set of columns from the set of columns of A.

We will now formalize the previous example, which will make it trivial to determine a linearly independent set of vectors that will span the column space of a matrix, and is constituted of just columns of A.

Theorem BCS

Basis of the Column Space

Suppose that A is an $m \times n$ matrix with columns $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n$, and B is a row-equivalent matrix in reduced row-echelon form with r nonzero rows. Let $D = \{d_1, d_2, d_3, \ldots, d_r\}$

be the set of column indices where B has leading 1's. Let $T = \{\mathbf{A}_{d_1}, \mathbf{A}_{d_2}, \mathbf{A}_{d_3}, \dots, \mathbf{A}_{d_r}\}$. Then

- 1. T is a linearly independent set.
- 2. $\mathcal{C}(A) = \langle T \rangle$.

Proof Definition CSM [560] describes the column space as the span of the set of columns of A. Theorem BS [376] tells us that we can reduce the set of vectors used in a span. If we apply Theorem BS [376] to $\mathcal{C}(A)$, we would collect the columns of A into a matrix (which would just be A again) and bring the matrix to reduced row-echelon form, which is the matrix B in the statement of the theorem. In this case, the conclusions of Theorem BS [376] applied to A, B and $\mathcal{C}(A)$ are exactly the conclusions we desire.

This is a nice result since it gives us a handful of vectors that describe the entire column space (through the span), and we believe this set is as small as possible because we cannot create any more relations of linear dependence to trim it down further. Furthermore, we defined the column space (Definition CSM [560]) as all linear combinations of the columns of the matrix, and the elements of the set T are still columns of the matrix (we won't be so lucky in the next two constructions of the column space).

Procedurally this theorem is extremely easy to apply. Row-reduce the original matrix, identify r columns with leading 1's in this reduced matrix, and grab the corresponding columns of the original matrix. But it is still important to study the proof of Theorem BS

[376] and its motivation in Example COV [370] which lie at the root of this theorem. We'll trot through an example all the same.

Example CSOCD

Column space, original columns, Archetype D

Let's determine a compact expression for the entire column space of the coefficient matrix of the system of equations that is Archetype D [1670]. Notice that in Example CSMCS [561] we were only determining if individual vectors were in the column space or not, now we are describing the entire column space.

To start with the application of Theorem BCS [569], call the coefficient matrix A

$$A = \begin{bmatrix} 2 & 1 & 7 & -7 \\ -3 & 4 & -5 & -6 \\ 1 & 1 & 4 & -5 \end{bmatrix}.$$

and row-reduce it to reduced row-echelon form,

$$B = \begin{bmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are leading 1's in columns 1 and 2, so $D = \{1, 2\}$. To construct a set that spans $\mathcal{C}(A)$, just grab the columns of A indicated by the set D, so

$$\mathcal{C}(A) = \left\langle \left\{ \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 1\\4\\1 \end{bmatrix} \right\} \right\rangle.$$

That's it.

In Example CSMCS [561] we determined that the vector

$$\mathbf{c} = \begin{bmatrix} 2\\3\\2 \end{bmatrix}$$

was not in the column space of A. Try to write **c** as a linear combination of the first two columns of A. What happens?

Also in Example CSMCS [561] we determined that the vector

$$\mathbf{b} = \begin{bmatrix} 8\\ -12\\ 4 \end{bmatrix}$$

was in the column space of A. Try to write **b** as a linear combination of the first two columns of A. What happens? Did you find a unique solution to this question? Hmmmm.

Subsection CSNM Column Space of a Nonsingular Matrix

Let's specialize to square matrices and contrast the column spaces of the coefficient matrices in Archetype A [1642] and Archetype B [1652].

Example CSAA Column space of Archetype A

The coefficient matrix in Archetype A [1642] is

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

which row-reduces to

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Columns 1 and 2 have leading 1's, so by Theorem BCS [569] we can write

$$\mathcal{C}(A) = \langle \{\mathbf{A}_1, \, \mathbf{A}_2\} \rangle = \left\langle \left\{ \begin{bmatrix} 1\\2\\1 \end{bmatrix}, \, \begin{bmatrix} -1\\1\\1 \end{bmatrix} \right\} \right\rangle.$$

We want to show in this example that $\mathcal{C}(A) \neq \mathbb{C}^3$. So take, for example, the vector $\mathbf{b} = \begin{bmatrix} 1\\ 3\\ 2 \end{bmatrix}$.

Then there is no solution to the system $\mathcal{LS}(A, \mathbf{b})$, or equivalently, it is not possible to write **b** as a linear combination of \mathbf{A}_1 and \mathbf{A}_2 . Try one of these two computations yourself. (Or try both!). Since $\mathbf{b} \notin \mathcal{C}(A)$, the column space of A cannot be all of \mathbb{C}^3 . So by varying the vector

of constants, it is possible to create inconsistent systems of equations with this coefficient matrix (the vector \mathbf{b} being one such example).

In Example MWIAA [506] we wished to show that the coefficient matrix from Archetype A [1642] was not invertible as a first example of a matrix without an inverse. Our device there was to find an inconsistent linear system with A as the coefficient matrix. The vector of constants in that example was **b**, deliberately chosen outside the column space of A.

Example CSAB Column space of Archetype B

The coefficient matrix in Archetype B [1652], call it *B* here, is known to be nonsingular (see Example NM [174]). By Theorem NMUS [180], the linear system $\mathcal{LS}(B, \mathbf{b})$ has a (unique) solution for every choice of **b**. Theorem CSCS [563] then says that $\mathbf{b} \in \mathcal{C}(B)$ for all $\mathbf{b} \in \mathbb{C}^3$. Stated differently, there is no way to build an inconsistent system with the coefficient matrix *B*, but then we knew that already from Theorem NMUS [180].

Example CSAA [573] and Example CSAB [574] together motivate the following equivalence, which says that nonsingular matrices have column spaces that are as big as possible.

Theorem CSNM

Column Space of a Nonsingular Matrix

Suppose A is a square matrix of size n. Then A is nonsingular if and only if $\mathcal{C}(A) = \mathbb{C}^n$. \Box

Proof (\Rightarrow) Suppose A is nonsingular. We wish to establish the set equality $\mathcal{C}(A) = \mathbb{C}^n$. By Definition CSM [560], $\mathcal{C}(A) \subseteq \mathbb{C}^n$. To show that $\mathbb{C}^n \subseteq \mathcal{C}(A)$ choose $\mathbf{b} \in \mathbb{C}^n$. By Theorem NMUS [180], we know the linear system $\mathcal{LS}(A, \mathbf{b})$ has a (unique) solution and therefore is consistent. Theorem CSCS [563] then says that $\mathbf{b} \in \mathcal{C}(A)$. So by Definition SE [1604], $\mathcal{C}(A) = \mathbb{C}^n$.

(\Leftarrow) If \mathbf{e}_i is column *i* of the $n \times n$ identity matrix (Definition SUV [409]) and by hypothesis $\mathcal{C}(A) = \mathbb{C}^n$, then $\mathbf{e}_i \in \mathcal{C}(A)$ for $1 \leq i \leq n$. By Theorem CSCS [563], the system $\mathcal{LS}(A, \mathbf{e}_i)$ is consistent for $1 \leq i \leq n$. Let \mathbf{b}_i denote any one particular solution to $\mathcal{LS}(A, \mathbf{e}_i)$, $1 \leq i \leq n$.

Define the $n \times n$ matrix $B = [\mathbf{b}_1 | \mathbf{b}_2 | \mathbf{b}_3 | \dots | \mathbf{b}_n]$. Then

$$AB = A [\mathbf{b}_1 | \mathbf{b}_2 | \mathbf{b}_3 | \dots | \mathbf{b}_n]$$

= $[A\mathbf{b}_1 | A\mathbf{b}_2 | A\mathbf{b}_3 | \dots | A\mathbf{b}_n]$ Definition MM [468]
= $[\mathbf{e}_1 | \mathbf{e}_2 | \mathbf{e}_3 | \dots | \mathbf{e}_n]$
= I_n Definition SUV [409]

So the matrix B is a "right-inverse" for A. By Theorem NMRRI [176], I_n is a nonsingular matrix, so by Theorem NPNT [538] both A and B are nonsingular. Thus, in particular, A is nonsingular. (Travis Osborne contributed to this proof.)

With this equivalence for nonsingular matrices we can update our list, Theorem NME3 [543].

Theorem NME4 Nonsingular Matrix Equivalences, Round 4

Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.
- 7. The column space of A is \mathbb{C}^n , $\mathcal{C}(A) = \mathbb{C}^n$.

Proof Since Theorem CSNM [574] is an equivalence, we can add it to the list in Theorem NME3 [543]. ■

Subsection RSM Row Space of a Matrix

The rows of a matrix can be viewed as vectors, since they are just lists of numbers, arranged horizontally. So we will transpose a matrix, turning rows into columns, so we can then

 \square

manipulate rows as column vectors. As a result we will be able to make some new connections between row operations and solutions to systems of equations. OK, here is the second primary definition of this section.

Definition RSM Row Space of a Matrix

Suppose A is an $m \times n$ matrix. Then the **row space** of A, $\mathcal{R}(A)$, is the column space of A^t , i.e. $\mathcal{R}(A) = \mathcal{C}(A^t)$.

(This definition contains Notation RSM.)

Informally, the row space is the set of all linear combinations of the rows of A. However, we write the rows as column vectors, thus the necessity of using the transpose to make the rows into columns. Additionally, with the row space defined in terms of the column space, all of the previous results of this section can be applied to row spaces.

Notice that if A is a rectangular $m \times n$ matrix, then $\mathcal{C}(A) \subseteq \mathbb{C}^m$, while $\mathcal{R}(A) \subseteq \mathbb{C}^n$ and the two sets are not comparable since they do not even hold objects of the same type. However, when A is square of size n, both $\mathcal{C}(A)$ and $\mathcal{R}(A)$ are subsets of \mathbb{C}^n , though usually the sets will not be equal (but see Exercise CRS.M20 [599]).

Example RSAI Row space of Archetype I \triangle

The coefficient matrix in Archetype I [1714] is

$$I = \begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 \\ 0 & 0 & 2 & -3 & -4 & 12 & -8 \\ -1 & -4 & 2 & 4 & 8 & -31 & 37 \end{bmatrix}.$$

To build the row space, we transpose the matrix,

$$I^{t} = \begin{bmatrix} 1 & 2 & 0 & -1 \\ 4 & 8 & 0 & -4 \\ 0 & -1 & 2 & 2 \\ -1 & 3 & -3 & 4 \\ 0 & 9 & -4 & 8 \\ 7 & -13 & 12 & -31 \\ -9 & 7 & -8 & 37 \end{bmatrix}$$

Then the columns of this matrix are used in a span to build the row space,

$$\mathcal{R}(I) = \mathcal{C}(I^{t}) = \left\langle \left\{ \begin{bmatrix} 1\\4\\0\\-1\\0\\7\\-9 \end{bmatrix}, \begin{bmatrix} 2\\8\\-1\\3\\9\\-13\\7 \end{bmatrix}, \begin{bmatrix} 0\\0\\2\\-3\\-4\\12\\-8 \end{bmatrix}, \begin{bmatrix} -1\\-4\\2\\4\\8\\-31\\37 \end{bmatrix} \right\} \right\rangle.$$

However, we can use Theorem BCS [569] to get a slightly better description. First, row-reduce I^t ,

Since there are leading 1's in columns with indices $D = \{1, 2, 3\}$, the column space of I^t can be spanned by just the first three columns of I^t ,

$$\mathcal{R}(I) = \mathcal{C}(I^{t}) = \left\langle \left\{ \begin{bmatrix} 1\\4\\0\\-1\\0\\7\\-9 \end{bmatrix}, \begin{bmatrix} 2\\8\\-1\\3\\9\\-13\\7 \end{bmatrix}, \begin{bmatrix} 0\\0\\2\\-3\\-4\\12\\-8 \end{bmatrix} \right\} \right\rangle.$$

 \boxtimes

The row space would not be too interesting if it was simply the column space of the transpose. However, when we do row operations on a matrix we have no effect on the many

linear combinations that can be formed with the rows of the matrix. This is stated more carefully in the following theorem.

Theorem REMRS Row-Equivalent Matrices have equal Row Spaces Suppose A and B are row-equivalent matrices. Then $\mathcal{R}(A) = \mathcal{R}(B)$.

Proof Two matrices are row-equivalent (Definition REM [60]) if one can be obtained from another by a sequence of (possibly many) row operations. We will prove the theorem for two matrices that differ by a single row operation, and then this result can be applied repeatedly to get the full statement of the theorem. The row spaces of A and B are spans of the columns of their transposes. For each row operation we perform on a matrix, we can define an analogous operation on the columns. Perhaps we should call these **column operations**. Instead, we will still call them row operations, but we will apply them to the columns of the transposes.

Refer to the columns of A^t and B^t as \mathbf{A}_i and \mathbf{B}_i , $1 \leq i \leq m$. The row operation that switches rows will just switch columns of the transposed matrices. This will have no effect on the possible linear combinations formed by the columns.

Suppose that B^t is formed from A^t by multiplying column \mathbf{A}_t by $\alpha \neq 0$. In other words, $\mathbf{B}_t = \alpha \mathbf{A}_t$, and $\mathbf{B}_i = \mathbf{A}_i$ for all $i \neq t$. We need to establish that two sets are equal, $\mathcal{C}(A^t) = \mathcal{C}(B^t)$. We will take a generic element of one and show that it is contained in the other.

$$\beta_1 \mathbf{B}_1 + \beta_2 \mathbf{B}_2 + \beta_3 \mathbf{B}_3 + \dots + \beta_t \mathbf{B}_t + \dots + \beta_m \mathbf{B}_m$$

 \square

$$= \beta_1 \mathbf{A}_1 + \beta_2 \mathbf{A}_2 + \beta_3 \mathbf{A}_3 + \dots + \beta_t (\alpha \mathbf{A}_t) + \dots + \beta_m \mathbf{A}_m$$
$$= \beta_1 \mathbf{A}_1 + \beta_2 \mathbf{A}_2 + \beta_3 \mathbf{A}_3 + \dots + (\alpha \beta_t) \mathbf{A}_t + \dots + \beta_m \mathbf{A}_m$$

says that $\mathcal{C}(B^t) \subseteq \mathcal{C}(A^t)$. Similarly,

$$\gamma_{1}\mathbf{A}_{1}+\gamma_{2}\mathbf{A}_{2}+\gamma_{3}\mathbf{A}_{3}+\cdots+\gamma_{t}\mathbf{A}_{t}+\cdots+\gamma_{m}\mathbf{A}_{m}$$

$$=\gamma_{1}\mathbf{A}_{1}+\gamma_{2}\mathbf{A}_{2}+\gamma_{3}\mathbf{A}_{3}+\cdots+\left(\frac{\gamma_{t}}{\alpha}\alpha\right)\mathbf{A}_{t}+\cdots+\gamma_{m}\mathbf{A}_{m}$$

$$=\gamma_{1}\mathbf{A}_{1}+\gamma_{2}\mathbf{A}_{2}+\gamma_{3}\mathbf{A}_{3}+\cdots+\frac{\gamma_{t}}{\alpha}\left(\alpha\mathbf{A}_{t}\right)+\cdots+\gamma_{m}\mathbf{A}_{m}$$

$$=\gamma_{1}\mathbf{B}_{1}+\gamma_{2}\mathbf{B}_{2}+\gamma_{3}\mathbf{B}_{3}+\cdots+\frac{\gamma_{t}}{\alpha}\mathbf{B}_{t}+\cdots+\gamma_{m}\mathbf{B}_{m}$$

says that $\mathcal{C}(A^t) \subseteq \mathcal{C}(B^t)$. So $\mathcal{R}(A) = \mathcal{C}(A^t) = \mathcal{C}(B^t) = \mathcal{R}(B)$ when a single row operation of the second type is performed.

Suppose now that B^t is formed from A^t by replacing \mathbf{A}_t with $\alpha \mathbf{A}_s + \mathbf{A}_t$ for some $\alpha \in \mathbb{C}$ and $s \neq t$. In other words, $\mathbf{B}_t = \alpha \mathbf{A}_s + \mathbf{A}_t$, and $\mathbf{B}_i = \mathbf{A}_i$ for $i \neq t$.

$$\beta_{1}\mathbf{B}_{1}+\beta_{2}\mathbf{B}_{2}+\beta_{3}\mathbf{B}_{3}+\dots+\beta_{s}\mathbf{B}_{s}+\dots+\beta_{t}\mathbf{B}_{t}+\dots+\beta_{m}\mathbf{B}_{m}$$

$$=\beta_{1}\mathbf{A}_{1}+\beta_{2}\mathbf{A}_{2}+\beta_{3}\mathbf{A}_{3}+\dots+\beta_{s}\mathbf{A}_{s}+\dots+\beta_{t}\left(\alpha\mathbf{A}_{s}+\mathbf{A}_{t}\right)+\dots+\beta_{m}\mathbf{A}_{m}$$

$$=\beta_{1}\mathbf{A}_{1}+\beta_{2}\mathbf{A}_{2}+\beta_{3}\mathbf{A}_{3}+\dots+\beta_{s}\mathbf{A}_{s}+\dots+\left(\beta_{t}\alpha\right)\mathbf{A}_{s}+\beta_{t}\mathbf{A}_{t}+\dots+\beta_{m}\mathbf{A}_{m}$$

$$=\beta_{1}\mathbf{A}_{1}+\beta_{2}\mathbf{A}_{2}+\beta_{3}\mathbf{A}_{3}+\dots+\beta_{s}\mathbf{A}_{s}+\left(\beta_{t}\alpha\right)\mathbf{A}_{s}+\dots+\beta_{t}\mathbf{A}_{t}+\dots+\beta_{m}\mathbf{A}_{m}$$

$$=\beta_{1}\mathbf{A}_{1}+\beta_{2}\mathbf{A}_{2}+\beta_{3}\mathbf{A}_{3}+\dots+\left(\beta_{s}+\beta_{t}\alpha\right)\mathbf{A}_{s}+\dots+\beta_{t}\mathbf{A}_{t}+\dots+\beta_{m}\mathbf{A}_{m}$$

says that $\mathcal{C}(B^t) \subseteq \mathcal{C}(A^t)$. Similarly,

$$\gamma_{1}\mathbf{A}_{1} + \gamma_{2}\mathbf{A}_{2} + \gamma_{3}\mathbf{A}_{3} + \dots + \gamma_{s}\mathbf{A}_{s} + \dots + \gamma_{t}\mathbf{A}_{t} + \dots + \gamma_{m}\mathbf{A}_{m}$$

$$= \gamma_{1}\mathbf{A}_{1} + \gamma_{2}\mathbf{A}_{2} + \gamma_{3}\mathbf{A}_{3} + \dots + \gamma_{s}\mathbf{A}_{s} + \dots + (-\alpha\gamma_{t}\mathbf{A}_{s} + \alpha\gamma_{t}\mathbf{A}_{s}) + \gamma_{t}\mathbf{A}_{t} + \dots + \gamma_{m}\mathbf{A}_{m}$$

$$= \gamma_{1}\mathbf{A}_{1} + \gamma_{2}\mathbf{A}_{2} + \gamma_{3}\mathbf{A}_{3} + \dots + (-\alpha\gamma_{t}\mathbf{A}_{s}) + \gamma_{s}\mathbf{A}_{s} + \dots + (\alpha\gamma_{t}\mathbf{A}_{s} + \gamma_{t}\mathbf{A}_{t}) + \dots + \gamma_{m}\mathbf{A}_{m}$$

$$= \gamma_{1}\mathbf{A}_{1} + \gamma_{2}\mathbf{A}_{2} + \gamma_{3}\mathbf{A}_{3} + \dots + (-\alpha\gamma_{t} + \gamma_{s})\mathbf{A}_{s} + \dots + \gamma_{t}(\alpha\mathbf{A}_{s} + \mathbf{A}_{t}) + \dots + \gamma_{m}\mathbf{A}_{m}$$

$$= \gamma_{1}\mathbf{B}_{1} + \gamma_{2}\mathbf{B}_{2} + \gamma_{3}\mathbf{B}_{3} + \dots + (-\alpha\gamma_{t} + \gamma_{s})\mathbf{B}_{s} + \dots + \gamma_{t}\mathbf{B}_{t} + \dots + \gamma_{m}\mathbf{B}_{m}$$

says that $\mathcal{C}(A^t) \subseteq \mathcal{C}(B^t)$. So $\mathcal{R}(A) = \mathcal{C}(A^t) = \mathcal{C}(B^t) = \mathcal{R}(B)$ when a single row operation of the third type is performed.

So the row space of a matrix is preserved by each row operation, and hence row spaces of row-equivalent matrices are equal sets.

Example RSREM

Row spaces of two row-equivalent matrices

In Example TREM [60] we saw that the matrices

$$A = \begin{bmatrix} 2 & -1 & 3 & 4 \\ 5 & 2 & -2 & 3 \\ 1 & 1 & 0 & 6 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 1 & 1 & 0 & 6 \\ 3 & 0 & -2 & -9 \\ 2 & -1 & 3 & 4 \end{bmatrix}$$

are row-equivalent by demonstrating a sequence of two row operations that converted A into

B. Applying Theorem REMRS [580] we can say

$$\mathcal{R}(A) = \left\langle \left\{ \begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix}, \begin{bmatrix} 5\\2\\-2\\3 \end{bmatrix}, \begin{bmatrix} 1\\1\\0\\6 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 1\\1\\0\\6 \end{bmatrix}, \begin{bmatrix} 3\\0\\-2\\-9 \end{bmatrix}, \begin{bmatrix} 2\\-1\\3\\4 \end{bmatrix} \right\} \right\rangle = \mathcal{R}(B)$$

 \boxtimes

Theorem REMRS [580] is at its best when one of the row-equivalent matrices is in reduced row-echelon form. The vectors that correspond to the zero rows can be ignored. (Who needs the zero vector when building a span? See Exercise LI.T10 [346].) The echelon pattern insures that the nonzero rows yield vectors that are linearly independent. Here's the theorem.

Theorem BRS Basis for the Row Space

Suppose that A is a matrix and B is a row-equivalent matrix in reduced row-echelon form. Let S be the set of nonzero columns of B^t . Then

1. $\mathcal{R}(A) = \langle S \rangle$.

2. S is a linearly independent set.

Proof From Theorem REMRS [580] we know that $\mathcal{R}(A) = \mathcal{R}(B)$. If *B* has any zero rows, these correspond to columns of B^t that are the zero vector. We can safely toss out the zero vector in the span construction, since it can be recreated from the nonzero vectors by a linear combination where all the scalars are zero. So $\mathcal{R}(A) = \langle S \rangle$.

Suppose *B* has *r* nonzero rows and let $D = \{d_1, d_2, d_3, \ldots, d_r\}$ denote the column indices of *B* that have a leading one in them. Denote the *r* column vectors of B^t , the vectors in *S*, as $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \ldots, \mathbf{B}_r$. To show that *S* is linearly independent, start with a relation of linear dependence

$$\alpha_1 \mathbf{B}_1 + \alpha_2 \mathbf{B}_2 + \alpha_3 \mathbf{B}_3 + \dots + \alpha_r \mathbf{B}_r = \mathbf{0}$$

Now consider this vector equality in location d_i . Since B is in reduced row-echelon form, the entries of column d_i of B are all zero, except for a (leading) 1 in row i. Thus, in B^t , row d_i is all zeros, excepting a 1 in column i. So, for $1 \le i \le r$,

$$0 = [\mathbf{0}]_{d_i}$$

$$= [\alpha_1 \mathbf{B}_1 + \alpha_2 \mathbf{B}_2 + \alpha_3 \mathbf{B}_3 + \dots + \alpha_r \mathbf{B}_r]_{d_i}$$
Definition RLDCV [53]
$$= [\alpha_1 \mathbf{B}_1]_{d_i} + [\alpha_2 \mathbf{B}_2]_{d_i} + [\alpha_3 \mathbf{B}_3]_{d_i} + \dots + [\alpha_r \mathbf{B}_r]_{d_i} +$$
Definition MA [427]
$$= \alpha_1 [\mathbf{B}_1]_{d_i} + \alpha_2 [\mathbf{B}_2]_{d_i} + \alpha_3 [\mathbf{B}_3]_{d_i} + \dots + \alpha_r [\mathbf{B}_r]_{d_i} +$$
Definition MSM [428]
$$= \alpha_1(0) + \alpha_2(0) + \alpha_3(0) + \dots + \alpha_i(1) + \dots + \alpha_r(0)$$
Definition RREF [63]
$$= \alpha_i$$

So we conclude that $\alpha_i = 0$ for all $1 \leq i \leq r$, establishing the linear independence of S

(Definition LICV [317]).

Example IAS Improving a span

Suppose in the course of analyzing a matrix (its column space, its null space, its...) we encounter the following set of vectors, described by a span

$$X = \left\langle \left\{ \begin{bmatrix} 1\\2\\1\\6\\6 \end{bmatrix}, \begin{bmatrix} 3\\-1\\2\\-1\\6 \end{bmatrix}, \begin{bmatrix} 1\\-1\\0\\-1\\-2 \end{bmatrix}, \begin{bmatrix} -3\\2\\-3\\6\\-10 \end{bmatrix} \right\} \right\rangle$$

Let A be the matrix whose rows are the vectors in X, so by design $X = \mathcal{R}(A)$,

$$A = \begin{bmatrix} 1 & 2 & 1 & 6 & 6 \\ 3 & -1 & 2 & -1 & 6 \\ 1 & -1 & 0 & -1 & -2 \\ -3 & 2 & -3 & 6 & -10 \end{bmatrix}$$

Row-reduce A to form a row-equivalent matrix in reduced row-echelon form,

$$B = \begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & -2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Then Theorem BRS [583] says we can grab the nonzero columns of B^t and write

$$X = \mathcal{R}(A) = \mathcal{R}(B) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0\\2\\-1 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\3\\1 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\-2\\5 \end{bmatrix} \right\} \right\rangle$$

These three vectors provide a much-improved description of X. There are fewer vectors, and the pattern of zeros and ones in the first three entries makes it easier to determine membership in X. And all we had to do was row-reduce the right matrix and toss out a zero row. Next to row operations themselves, *this is probably the most powerful computational technique at your disposal* as it quickly provides a much improved description of a span, any span.

Theorem BRS [583] and the techniques of Example IAS [585] will provide yet another description of the column space of a matrix. First we state a triviality as a theorem, so we can reference it later.

Theorem CSRST Column Space, Row Space, Transpose Suppose A is a matrix. Then $C(A) = \mathcal{R}(A^t)$.

Proof

$$\mathcal{C}(A) = \mathcal{C}\left(\left(A^{t}\right)^{t}\right)$$
 Theorem TT [437]

 $= \mathcal{R}(A^t)$ Definition RSM [577]

So to find another expression for the column space of a matrix, build its transpose, row-reduce it, toss out the zero rows, and convert the nonzero rows to column vectors to yield an improved set for the span construction. We'll do Archetype I [1714], then you do Archetype J [1723].

Example CSROI

Column space from row operations, Archetype I

To find the column space of the coefficient matrix of Archetype I [1714], we proceed as follows. The matrix is

$$I = \begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 \\ 0 & 0 & 2 & -3 & -4 & 12 & -8 \\ -1 & -4 & 2 & 4 & 8 & -31 & 37 \end{bmatrix}.$$

The transpose is

$$\begin{bmatrix} 1 & 2 & 0 & -1 \\ 4 & 8 & 0 & -4 \\ 0 & -1 & 2 & 2 \\ -1 & 3 & -3 & 4 \\ 0 & 9 & -4 & 8 \\ 7 & -13 & 12 & -31 \\ -9 & 7 & -8 & 37 \end{bmatrix}.$$

Row-reduced this becomes,

Now, using Theorem CSRST [586] and Theorem BRS [583]

$$\mathcal{C}(I) = \mathcal{R}(I^{t}) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0\\-\frac{31}{7} \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\\frac{12}{7} \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\\frac{13}{7} \end{bmatrix} \right\} \right\rangle.$$

This is a very nice description of the column space. Fewer vectors than the 7 involved in the definition, and the pattern of the zeros and ones in the first 3 slots can be used to advantage. For example, Archetype I [1714] is presented as a consistent system of equations with a vector of constants

$$\mathbf{b} = \begin{bmatrix} 3\\9\\1\\4 \end{bmatrix}$$

Since $\mathcal{LS}(I, \mathbf{b})$ is consistent, Theorem CSCS [563] tells us that $\mathbf{b} \in \mathcal{C}(I)$. But we could see this quickly with the following computation, which really only involves any work in the 4th entry of the vectors as the scalars in the linear combination are *dictated* by the first three entries of **b**.

$$\mathbf{b} = \begin{bmatrix} 3\\9\\1\\4 \end{bmatrix} = 3\begin{bmatrix} 1\\0\\0\\-\frac{31}{7} \end{bmatrix} + 9\begin{bmatrix} 0\\1\\0\\\frac{12}{7} \end{bmatrix} + 1\begin{bmatrix} 0\\0\\1\\\frac{13}{7} \end{bmatrix}$$

Can you now rapidly construct several vectors, **b**, so that $\mathcal{LS}(I, \mathbf{b})$ is consistent, and several more so that the system is inconsistent?

Subsection READ Reading Questions

1. Write the column space of the matrix below as the span of a set of three vectors and explain your choice of method.

$$\begin{bmatrix} 1 & 3 & 1 & 3 \\ 2 & 0 & 1 & 1 \\ -1 & 2 & 1 & 0 \end{bmatrix}$$

2. Suppose that A is an $n \times n$ nonsingular matrix. What can you say about its column space?

3. Is the vector $\begin{bmatrix} 0\\5\\2\\3 \end{bmatrix}$ in the row space of the following matrix? Why or why not?

$$\begin{bmatrix} 1 & 3 & 1 & 3 \\ 2 & 0 & 1 & 1 \\ -1 & 2 & 1 & 0 \end{bmatrix}$$

Subsection EXC Exercises

C20 For parts (a), (b) and c, find a set of linearly independent vectors X so that $\mathcal{C}(A) = \langle X \rangle$, and a set of linearly independent vectors Y so that $\mathcal{R}(A) = \langle Y \rangle$.

a).
$$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 2 \\ 1 & -1 & 2 & 3 \\ 1 & 1 & 2 & -1 \end{bmatrix}$$

b).
$$A = \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 3 & 2 & -1 & 4 & 5 \\ 0 & 1 & 1 & 1 & 2 \end{bmatrix}$$

c).
$$A = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 0 & 3 \\ 1 & 2 & -3 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \end{bmatrix}$$

d). From your results in parts (a) - (c), can you formulate a conjecture about the sets X and Y?

Contributed by Chris Black

C30 Example CSOCD [571] expresses the column space of the coefficient matrix from Archetype D [1670] (call the matrix A here) as the span of the first two columns of A. In Example CSMCS [561] we determined that the vector

$$\mathbf{c} = \begin{bmatrix} 2\\ 3\\ 2 \end{bmatrix}$$

was not in the column space of A and that the vector

$$\mathbf{b} = \begin{bmatrix} 8\\-12\\4 \end{bmatrix}$$

was in the column space of A. Attempt to write **c** and **b** as linear combinations of the two vectors in the span construction for the column space in Example CSOCD [571] and record your observations.

Contributed by Robert Beezer Solution [601]

C31 For the matrix A below find a set of vectors T meeting the following requirements: (1) the span of T is the column space of A, that is, $\langle T \rangle = C(A)$, (2) T is linearly independent,

and (3) the elements of T are columns of A.

$$A = \begin{bmatrix} 2 & 1 & 4 & -1 & 2 \\ 1 & -1 & 5 & 1 & 1 \\ -1 & 2 & -7 & 0 & 1 \\ 2 & -1 & 8 & -1 & 2 \end{bmatrix}$$

Contributed by Robert Beezer Solution [601]

C32 In Example CSAA [573], verify that the vector **b** is not in the column space of the coefficient matrix. Contributed by Robert Beezer

C33 Find a linearly independent set S so that the span of S, $\langle S \rangle$, is row space of the matrix B, and S is linearly independent.

$$B = \begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ -1 & 2 & 3 & -4 \end{bmatrix}$$

Contributed by Robert Beezer Solution [602]

C34 For the 3×4 matrix A and the column vector $\mathbf{y} \in \mathbb{C}^4$ given below, determine if \mathbf{y} is in the row space of A. In other words, answer the question: $\mathbf{y} \in \mathcal{R}(A)$?

$$A = \begin{bmatrix} -2 & 6 & 7 & -1 \\ 7 & -3 & 0 & -3 \\ 8 & 0 & 7 & 6 \end{bmatrix} \qquad \qquad \mathbf{y} = \begin{bmatrix} 2 \\ 1 \\ 3 \\ -2 \end{bmatrix}$$

Contributed by Robert Beezer Solution [603]

C35 For the matrix A below, find two different linearly independent sets whose spans equal the column space of A, C(A), such that

(a) the elements are each columns of A.

(b) the set is obtained by a procedure that is substantially different from the procedure you use in part (a).

$$A = \begin{bmatrix} 3 & 5 & 1 & -2 \\ 1 & 2 & 3 & 3 \\ -3 & -4 & 7 & 13 \end{bmatrix}$$

Contributed by Robert Beezer Solution [603]

C40 The following archetypes are systems of equations. For each system, write the vector of constants as a linear combination of the vectors in the span construction for the column

space provided by Theorem BCS [569] (these vectors are listed for each of these archetypes). Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670] Archetype E [1678] Archetype F [1686] Archetype G [1697] Archetype H [1705] Archetype I [1714] Archetype J [1723]

Contributed by Robert Beezer

C42 The following archetypes are either matrices or systems of equations with coefficient matrices. For each matrix, compute a set of column vectors such that (1) the vectors are columns of the matrix, (2) the set is linearly independent, and (3) the span of the set is the column space of the matrix. See Theorem BCS [569].

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678] Archetype F [1686] Archetype G [1697]/Archetype H [1705] Archetype I [1714] Archetype J [1723] Archetype K [1734] Archetype L [1742]

Contributed by Robert Beezer

C50 The following archetypes are either matrices or systems of equations with coefficient matrices. For each matrix, compute a set of column vectors such that (1) the set is linearly independent, and (2) the span of the set is the row space of the matrix. See Theorem BRS [583].

```
Archetype A [1642]
Archetype B [1652]
Archetype C [1662]
Archetype D [1670]/Archetype E [1678]
Archetype F [1686]
Archetype G [1697]/Archetype H [1705]
Archetype I [1714]
Archetype J [1723]
Archetype K [1734]
Archetype L [1742]
```

Contributed by Robert Beezer

C51 The following archetypes are either matrices or systems of equations with coefficient matrices. For each matrix, compute the column space as the span of a linearly independent set as follows: transpose the matrix, row-reduce, toss out zero rows, convert rows into column vectors. See Example CSROI [587].

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678] Archetype F [1686] Archetype G [1697]/Archetype H [1705] Archetype I [1714] Archetype J [1723] Archetype K [1734] Archetype L [1742]

Contributed by Robert Beezer

C52 The following archetypes are systems of equations. For each different coefficient matrix build two new vectors of constants. The first should lead to a consistent system and the

second should lead to an inconsistent system. Descriptions of the column space as spans of linearly independent sets of vectors with "nice patterns" of zeros and ones might be most useful and instructive in connection with this exercise. (See the end of Example CSROI [587].)

```
Archetype A [1642]
Archetype B [1652]
Archetype C [1662]
Archetype D [1670]/Archetype E [1678]
Archetype F [1686]
Archetype G [1697]/Archetype H [1705]
Archetype I [1714]
Archetype J [1723]
```

Contributed by Robert Beezer

M10 For the matrix E below, find vectors \mathbf{b} and \mathbf{c} so that the system $\mathcal{LS}(E, \mathbf{b})$ is consistent and $\mathcal{LS}(E, \mathbf{c})$ is inconsistent.

$$E = \begin{bmatrix} -2 & 1 & 1 & 0\\ 3 & -1 & 0 & 2\\ 4 & 1 & 1 & 6 \end{bmatrix}$$

Contributed by Robert Beezer Solution [605]

M20 Usually the column space and null space of a matrix contain vectors of different sizes. For a square matrix, though, the vectors in these two sets are the same size. Usually the two sets will be different. Construct an example of a square matrix where the column space and null space are equal.

Contributed by Robert Beezer Solution [605]

M21 We have a variety of theorems about how to create column spaces and row spaces and they frequently involve row-reducing a matrix. Here is a procedure that some try to use to get a column space. Begin with an $m \times n$ matrix A and row-reduce to a matrix B with columns $\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \ldots, \mathbf{B}_n$. Then form the column space of A as

 $\mathcal{C}(A) = \langle \{\mathbf{B}_1, \mathbf{B}_2, \mathbf{B}_3, \dots, \mathbf{B}_n\} \rangle = \mathcal{C}(B)$

This is *not* not a legitimate procedure, and therefore is *not* a theorem. Construct an example to show that the procedure will not in general create the column space of A. Contributed by Robert Beezer Solution [605]

T40 Suppose that A is an $m \times n$ matrix and B is an $n \times p$ matrix. Prove that the column space of AB is a subset of the column space of A, that is $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$. Provide an example where the opposite is false, in other words give an example where $\mathcal{C}(A) \not\subseteq \mathcal{C}(AB)$. (Compare with Exercise MM.T40 [494].) Contributed by Robert Beezer Solution [606]

T41 Suppose that A is an $m \times n$ matrix and B is an $n \times n$ nonsingular matrix. Prove that the column space of A is equal to the column space of AB, that is C(A) = C(AB). (Compare

with Exercise MM.T41 [494] and Exercise CRS.T40 [599].) Contributed by Robert Beezer Solution [606]

T45 Suppose that A is an $m \times n$ matrix and B is an $n \times m$ matrix where AB is a nonsingular matrix. Prove that

- (1) $\mathcal{N}(B) = \{\mathbf{0}\}$
- (2) $\mathcal{C}(B) \cap \mathcal{N}(A) = \{\mathbf{0}\}$

Discuss the case when m = n in connection with Theorem NPNT [538].

Contributed by Robert Beezer Solution [607]

Subsection SOL Solutions

C30 Contributed by Robert Beezer Statement [592]

In each case, begin with a vector equation where one side contains a linear combination of the two vectors from the span construction that gives the column space of A with unknowns for scalars, and then use Theorem SLSLC [229] to set up a system of equations. For \mathbf{c} , the corresponding system has no solution, as we would expect.

For **b** there is a solution, as we would expect. What is interesting is that the solution is unique. This is a consequence of the linear independence of the set of two vectors in the span construction. If we wrote **b** as a linear combination of all four columns of A, then there would be infinitely many ways to do this.

C31 Contributed by Robert Beezer Statement [592]

Theorem BCS [569] is the right tool for this problem. Row-reduce this matrix, identify the pivot columns and then grab the corresponding columns of A for the set T. The matrix A row-reduces to

$$\begin{bmatrix} 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

So $D = \{1, 2, 4, 5\}$ and then

$$T = \{\mathbf{A}_{1}, \, \mathbf{A}_{2}, \, \mathbf{A}_{4}, \, \mathbf{A}_{5}\} = \left\{ \begin{bmatrix} 2\\1\\-1\\2 \end{bmatrix}, \, \begin{bmatrix} 1\\-1\\2\\-1 \end{bmatrix}, \, \begin{bmatrix} -1\\1\\0\\-1 \end{bmatrix}, \, \begin{bmatrix} 2\\1\\1\\2 \end{bmatrix} \right\}$$

has the requested properties.

C33 Contributed by Robert Beezer Statement [593] Theorem BRS [583] is the most direct route to a set with these properties. Row-reduce, toss zero rows, keep the others. You could also transpose the matrix, then look for the column space by row-reducing the transpose and applying Theorem BCS [569]. We'll do the former,

$$B \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So the set S is

$$S = \left\{ \begin{bmatrix} 1\\0\\-1\\2 \end{bmatrix}, \begin{bmatrix} 0\\1\\1\\-1 \end{bmatrix} \right\}$$

C34 Contributed by Robert Beezer Statement [594]

$$\mathbf{y} \in \mathcal{R}(A) \iff \mathbf{y} \in \mathcal{C}(A^t)$$
$$\iff \mathcal{LS}(A^t, \mathbf{y}) \text{ is consistent}$$

Definition RSM [577] Theorem CSCS [563]

The augmented matrix $[A^t | \mathbf{y}]$ row reduces to

$\lceil 1 \rceil$	0	0	0]
0	1	0	0
0	0	1	0
0	0	0	1

and with a leading 1 in the final column Theorem RCLS [120] tells us the linear system is inconsistent and so $\mathbf{y} \notin \mathcal{R}(A)$.

C35 Contributed by Robert Beezer Statement [594]

(a) By Theorem BCS [569] we can row-reduce A, identify pivot columns with the set D, and "keep" those columns of A and we will have a set with the desired properties.

$$A \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -13 & -19 \\ 0 & 1 & 8 & 11 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So we have the set of pivot columns $D = \{1, 2\}$ and we "keep" the first two columns of A,

$$\left\{ \begin{bmatrix} 3\\1\\-3 \end{bmatrix}, \begin{bmatrix} 5\\2\\-4 \end{bmatrix} \right\}$$

(b) We can view the column space as the row space of the transpose (Theorem CSRST [586]). We can get a basis of the row space of a matrix quickly by bringing the matrix to reduced row-echelon form and keeping the nonzero rows as column vectors (Theorem BRS [583]). Here goes.

$$A^{t} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Taking the nonzero rows and tilting them up as columns gives us

$$\left\{ \begin{bmatrix} 1\\0\\-2 \end{bmatrix}, \begin{bmatrix} 0\\1\\3 \end{bmatrix} \right\}$$

An approach based on the matrix L from extended echelon form (Definition EEF [619]) and Theorem FS [625] will work as well as an alternative approach. M10 Contributed by Robert Beezer Statement [598]

Any vector from \mathbb{C}^3 will lead to a consistent system, and therefore there is no vector that will lead to an inconsistent system.

How do we convince ourselves of this? First, row-reduce E,

$$E \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

If we augment E with any vector of constants, and row-reduce the augmented matrix, we will never find a leading 1 in the final column, so by Theorem RCLS [120] the system will always be consistent.

Said another way, the column space of E is all of \mathbb{C}^3 , $\mathcal{C}(E) = \mathbb{C}^3$. So by Theorem CSCS [563] any vector of constants will create a consistent system (and none will create an inconsistent system).

M20Contributed by Robert BeezerStatement [599]The 2×2 matrix

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

has $\mathcal{C}(A) = \mathcal{N}(A) = \left\langle \left\{ \begin{bmatrix} 1\\ -1 \end{bmatrix} \right\} \right\rangle$.

M21 Contributed by Robert Beezer Statement [599] Begin with a matrix A (of any size) that does not have any zero rows, but which when row-reduced to B yields at least one row of zeros. Such a matrix should be easy to construct (or find, like say from Archetype A [1642]).

 $\mathcal{C}(A)$ will contain some vectors whose final slot (entry m) is non-zero, however, every column vector from the matrix B will have a zero in slot m and so every vector in $\mathcal{C}(B)$ will also contain a zero in the final slot. This means that $\mathcal{C}(A) \neq \mathcal{C}(B)$, since we have vectors in $\mathcal{C}(A)$ that cannot be elements of $\mathcal{C}(B)$.

T40 Contributed by Robert Beezer Statement [599] Choose $\mathbf{x} \in \mathcal{C}(AB)$. Then by Theorem CSCS [563] there is a vector \mathbf{w} that is a solution to $\mathcal{LS}(AB, \mathbf{x})$. Define the vector \mathbf{y} by $\mathbf{y} = B\mathbf{w}$. We're set,

$A\mathbf{y} = A\left(B\mathbf{w}\right)$	Definition of \mathbf{y}
$= (AB) \mathbf{w}$	Theorem MMA [478]
$= \mathbf{x}$	\mathbf{w} solution to $\mathcal{LS}(AB, \mathbf{x})$

This says that $\mathcal{LS}(A, \mathbf{x})$ is a consistent system, and by Theorem CSCS [563], we see that $\mathbf{x} \in \mathcal{C}(A)$ and therefore $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$.

For an example where $\mathcal{C}(A) \not\subseteq \mathcal{C}(AB)$ choose A to be any nonzero matrix and choose B to be a zero matrix. Then $\mathcal{C}(A) \neq \{\mathbf{0}\}$ and $\mathcal{C}(AB) = \mathcal{C}(\mathcal{O}) = \{\mathbf{0}\}.$

T41 Contributed by Robert Beezer Statement [599] From the solution to Exercise CRS.T40 [599] we know that $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$. So to establish the set equality (Definition SE [1604]) we need to show that $\mathcal{C}(A) \subseteq \mathcal{C}(AB)$. Choose $\mathbf{x} \in \mathcal{C}(A)$. By Theorem CSCS [563] the linear system $\mathcal{LS}(A, \mathbf{x})$ is consistent, so let \mathbf{y} be one such solution. Because B is nonsingular, and linear system using B as a coefficient matrix will have a solution (Theorem NMUS [180]). Let \mathbf{w} be the unique solution to the linear system $\mathcal{LS}(B, \mathbf{y})$. All set, here we go,

$(AB)\mathbf{w} = A\left(B\mathbf{w}\right)$	Theorem MMA $[478]$
$= A\mathbf{y}$	\mathbf{w} solution to $\mathcal{LS}(B, \mathbf{y})$
$= \mathbf{x}$	\mathbf{y} solution to $\mathcal{LS}(A, \mathbf{x})$

This says that the linear system $\mathcal{LS}(AB, \mathbf{x})$ is consistent, so by Theorem CSCS [563], $\mathbf{x} \in \mathcal{C}(AB)$. So $\mathcal{C}(A) \subseteq \mathcal{C}(AB)$.

T45 Contributed by Robert Beezer Statement [600] First, $\mathbf{0} \in \mathcal{N}(B)$ trivially. Now suppose that $\mathbf{x} \in \mathcal{N}(B)$. Then

$AB\mathbf{x} = A(B\mathbf{x})$	Theorem MMA $[478]$
= A 0	$\mathbf{x} \in \mathcal{N}(B)$
= 0	Theorem MMZM [474]

Since we have assumed AB is nonsingular, Definition NM [173] implies that $\mathbf{x} = \mathbf{0}$.

Second, $\mathbf{0} \in \mathcal{C}(B)$ and $\mathbf{0} \in \mathcal{N}(A)$ trivially, and so the zero vector is in the intersection as well (Definition SI [1608]). Now suppose that $\mathbf{y} \in \mathcal{C}(B) \cap \mathcal{N}(A)$. Because $\mathbf{y} \in \mathcal{C}(B)$, Theorem CSCS [563] says the system $\mathcal{LS}(B, \mathbf{y})$ is consistent. Let $\mathbf{x} \in \mathbb{C}^n$ be one solution to this system. Then

$AB\mathbf{x} = A(B\mathbf{x})$	Theorem MMA $[478]$	
$= A\mathbf{y}$	\mathbf{x} solution to $\mathcal{LS}(B, \mathbf{y})$	
= 0	$\mathbf{y} \in \mathcal{N}(A)$	

Since we have assumed AB is nonsingular, Definition NM [173] implies that $\mathbf{x} = \mathbf{0}$. Then $\mathbf{y} = B\mathbf{x} = B\mathbf{0} = \mathbf{0}$.

When AB is nonsingular and m = n we know that the first condition, $\mathcal{N}(B) = \{\mathbf{0}\}$, means that B is nonsingular (Theorem NMTNS [179]). Because B is nonsingular Theorem CSNM [574] implies that $\mathcal{C}(B) = \mathbb{C}^m$. In order to have the second condition fulfilled, $\mathcal{C}(B) \cap \mathcal{N}(A) = \{\mathbf{0}\}$, we must realize that $\mathcal{N}(A) = \{\mathbf{0}\}$. However, a second application of Theorem NMTNS [179] shows that A must be nonsingular. This reproduces Theorem NPNT [538].

Section FS Four Subsets

There are four natural subsets associated with a matrix. We have met three already: the null space, the column space and the row space. In this section we will introduce a fourth, the left null space. The objective of this section is to describe one procedure that will allow us to find linearly independent sets that span each of these four sets of column vectors. Along the way, we will make a connection with the inverse of a matrix, so Theorem FS [625] will tie together most all of this chapter (and the entire course so far).

Subsection LNS Left Null Space

Definition LNS Left Null Space Suppose A is an $m \times n$ matrix. Then the left null space is defined as $\mathcal{L}(A) = \mathcal{N}(A^t) \subseteq \mathbb{C}^m$. (This definition contains Notation LNS.)

The left null space will not feature prominently in the sequel, but we can explain its name and connect it to row operations. Suppose $\mathbf{y} \in \mathcal{L}(A)$. Then by Definition LNS [609], $A^t \mathbf{y} = \mathbf{0}$. We can then write

$0^t = \left(A^t \mathbf{y}\right)^t$	Definition LNS [609]
$=\mathbf{y}^{t}\left(A^{t}\right)^{t}$	Theorem MMT [482]
$= \mathbf{y}^t A$	Theorem TT $[437]$

The product $\mathbf{y}^t A$ can be viewed as the components of \mathbf{y} acting as the scalars in a linear combination of the *rows* of A. And the result is a "row vector", $\mathbf{0}^t$ that is totally zeros. When we apply a sequence of row operations to a matrix, each row of the resulting matrix is some linear combination of the rows. These observations tell us that the vectors in the left null space are scalars that record a sequence of row operations that result in a row of zeros in the row-reduced version of the matrix. We will see this idea more explicitly in the course of proving Theorem FS [625].

Example LNS Left null space We will find the left null space

We will find the left null space of

$$A = \begin{bmatrix} 1 & -3 & 1 \\ -2 & 1 & 1 \\ 1 & 5 & 1 \\ 9 & -4 & 0 \end{bmatrix}$$

 \triangle

We transpose A and row-reduce,

$$A^{t} = \begin{bmatrix} 1 & -2 & 1 & 9 \\ -3 & 1 & 5 & -4 \\ 1 & 1 & 1 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Applying Definition LNS [609] and Theorem BNS [334] we have

$$\mathcal{L}(A) = \mathcal{N}(A^t) = \left\langle \left\{ \begin{bmatrix} -2\\ 3\\ -1\\ 1 \end{bmatrix} \right\} \right\rangle$$

If you row-reduce A you will discover one zero row in the reduced row-echelon form. This zero row is created by a sequence of row operations, which in total amounts to a linear combination, with scalars $a_1 = -2$, $a_2 = 3$, $a_3 = -1$ and $a_4 = 1$, on the rows of A and which results in the zero vector (check this!). So the components of the vector describing the left null space of A provide a relation of linear dependence on the rows of A.

Subsection CCS Computing Column Spaces

We have three ways to build the column space of a matrix. First, we can use just the definition, Definition CSM [560], and express the column space as a span of the columns of the matrix. A second approach gives us the column space as the span of *some* of the columns of the matrix, but this set is linearly independent (Theorem BCS [569]). Finally, we can transpose the matrix, row-reduce the transpose, kick out zero rows, and transpose the remaining rows back into column vectors. Theorem CSRST [586] and Theorem BRS [583] tell us that the resulting vectors are linearly independent and their span is the column space of the original matrix.

We will now demonstrate a fourth method by way of a rather complicated example. Study this example carefully, but realize that its main purpose is to motivate a theorem that simplifies much of the apparent complexity. So other than an instructive exercise or two, the procedure we are about to describe will not be a usual approach to computing a column space.

Example CSANS Column space as null space Lets find the column space of the matrix A below with a new approach.

$$A = \begin{bmatrix} 10 & 0 & 3 & 8 & 7 \\ -16 & -1 & -4 & -10 & -13 \\ -6 & 1 & -3 & -6 & -6 \\ 0 & 2 & -2 & -3 & -2 \\ 3 & 0 & 1 & 2 & 3 \\ -1 & -1 & 1 & 1 & 0 \end{bmatrix}$$

By Theorem CSCS [563] we know that the column vector **b** is in the column space of A if and only if the linear system $\mathcal{LS}(A, \mathbf{b})$ is consistent. So let's try to solve this system in full generality, using a vector of variables for the vector of constants. In other words, which vectors **b** lead to consistent systems? Begin by forming the augmented matrix $[A \mid \mathbf{b}]$ with a general version of **b**,

$$[A \mid \mathbf{b}] = \begin{bmatrix} 10 & 0 & 3 & 8 & 7 & b_1 \\ -16 & -1 & -4 & -10 & -13 & b_2 \\ -6 & 1 & -3 & -6 & -6 & b_3 \\ 0 & 2 & -2 & -3 & -2 & b_4 \\ 3 & 0 & 1 & 2 & 3 & b_5 \\ -1 & -1 & 1 & 1 & 0 & b_6 \end{bmatrix}$$

To identify solutions we will row-reduce this matrix and bring it to reduced row-echelon form. Despite the presence of variables in the last column, there is nothing to stop us from doing this. Except our numerical routines on calculators can't be used, and even some of the symbolic algebra routines do some unexpected maneuvers with this computation. So do it by hand. Yes, it is a bit of work. But worth it. We'll still be here when you get back. Notice along the way that the row operations are *exactly* the same ones you would do if you were just row-reducing the coefficient matrix alone, say in connection with a homogeneous system of equations. The column with the b_i acts as a sort of bookkeeping device. There are many different possibilities for the result, depending on what order you choose to perform the row operations, but shortly we'll all be on the same page. Here's one possibility (you can find this same result by doing additional row operations with the fifth and sixth rows to remove any occurrences of b_1 and b_2 from the first four rows of your result):

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 2 & b_3 - b_4 + 2b_5 - b_6 \\ 0 & 1 & 0 & 0 & -3 & -2b_3 + 3b_4 - 3b_5 + 3b_6 \\ 0 & 0 & 1 & 0 & 1 & b_3 + b_4 + 3b_5 + 3b_6 \\ 0 & 0 & 0 & 1 & -2 & -2b_3 + b_4 - 4b_5 \\ 0 & 0 & 0 & 0 & b_1 + 3b_3 - b_4 + 3b_5 + b_6 \\ 0 & 0 & 0 & 0 & b_2 - 2b_3 + b_4 + b_5 - b_6 \end{bmatrix}$$

Our goal is to identify those vectors **b** which make $\mathcal{LS}(A, \mathbf{b})$ consistent. By Theorem RCLS [120] we know that the consistent systems are precisely those without a leading 1 in the last column. Are the expressions in the last column of rows 5 and 6 equal to zero, or are they leading 1's? The answer is: maybe. It depends on **b**. With a nonzero value for either of these expressions, we would scale the row and produce a leading 1. So we get a consistent system, and **b** is in the column space, if and only if these two expressions are both simultaneously

zero. In other words, members of the column space of A are exactly those vectors **b** that satisfy

$$b_1 + 3b_3 - b_4 + 3b_5 + b_6 = 0$$

$$b_2 - 2b_3 + b_4 + b_5 - b_6 = 0$$

Hmmm. Looks suspiciously like a homogeneous system of two equations with six variables. If you've been playing along (and we hope you have) then you may have a slightly different system, but you should have just two equations. Form the coefficient matrix and row-reduce (notice that the system above has a coefficient matrix that is already in reduced row-echelon form). We should all be together now with the same matrix,

$$L = \begin{bmatrix} 1 & 0 & 3 & -1 & 3 & 1 \\ 0 & 1 & -2 & 1 & 1 & -1 \end{bmatrix}$$

So, $C(A) = \mathcal{N}(L)$ and we can apply Theorem BNS [334] to obtain a linearly independent set to use in a span construction,

$$\mathcal{C}(A) = \mathcal{N}(L) = \left\langle \left\{ \begin{bmatrix} -3\\2\\1\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\-1\\0\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} -3\\-1\\0\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} -1\\1\\0\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} -1\\1\\0\\0\\1\\0 \end{bmatrix} \right\} \right\rangle$$

Whew! As a postscript to this central example, you may wish to convince yourself that the four vectors above really are elements of the column space? Do they create consistent systems with A as coefficient matrix? Can you recognize the constant vector in your description of these solution sets?

OK, that was so much fun, let's do it again. But simpler this time. And we'll all get the same results all the way through. Doing row operations by hand with variables can be a bit error prone, so let's see if we can improve the process some. Rather than row-reduce a column vector **b** full of variables, let's write $\mathbf{b} = I_6 \mathbf{b}$ and we will row-reduce the matrix I_6 and when we finish row-reducing, then we will compute the matrix-vector product. You should first convince yourself that we can operate like this (this is the subject of a future homework exercise). Rather than augmenting A with **b**, we will instead augment it with I_6 (does this feel familiar?),

$$M = \begin{bmatrix} 10 & 0 & 3 & 8 & 7 & 1 & 0 & 0 & 0 & 0 \\ -16 & -1 & -4 & -10 & -13 & 0 & 1 & 0 & 0 & 0 & 0 \\ -6 & 1 & -3 & -6 & -6 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & -2 & -3 & -2 & 0 & 0 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We want to row-reduce the left-hand side of this matrix, but we will apply the same row operations to the right-hand side as well. And once we get the left-hand side in reduced row-echelon form, we will continue on to put leading 1's in the final two rows, as well as clearing out the columns containing those two additional leading 1's. It is these additional row operations that will ensure that we all get to the same place, since the reduced rowechelon form is unique (Theorem RREFU [70]),

$$N = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 0 & 0 & 1 & -1 & 2 & -1 \\ 0 & 1 & 0 & 0 & -3 & 0 & 0 & -2 & 3 & -3 & 3 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 3 & 3 \\ 0 & 0 & 0 & 1 & -2 & 0 & 0 & -2 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 & -1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 & 1 & -1 \end{bmatrix}$$

We are after the final six columns of this matrix, which we will multiply by \mathbf{b}

$$J = \begin{bmatrix} 0 & 0 & 1 & -1 & 2 & -1 \\ 0 & 0 & -2 & 3 & -3 & 3 \\ 0 & 0 & 1 & 1 & 3 & 3 \\ 0 & 0 & -2 & 1 & -4 & 0 \\ 1 & 0 & 3 & -1 & 3 & 1 \\ 0 & 1 & -2 & 1 & 1 & -1 \end{bmatrix}$$
$$J\mathbf{b} = \begin{bmatrix} 0 & 0 & 1 & -1 & 2 & -1 \\ 0 & 0 & -2 & 3 & -3 & 3 \\ 0 & 0 & 1 & 1 & 3 & 3 \\ 0 & 0 & -2 & 1 & -4 & 0 \\ 1 & 0 & 3 & -1 & 3 & 1 \\ 0 & 1 & -2 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix} = \begin{bmatrix} b_3 - b_4 + 2b_5 - b_6 \\ -2b_3 + 3b_4 - 3b_5 + 3b_6 \\ -2b_3 + b_4 - 3b_5 + 3b_6 \\ -2b_3 + b_4 - 4b_5 \\ b_1 + 3b_3 - b_4 + 3b_5 + b_6 \\ b_2 - 2b_3 + b_4 + b_5 - b_6 \end{bmatrix}$$

SO

So by applying the same row operations that row-reduce A to the identity matrix (which we could do with a calculator once I_6 is placed alongside of A), we can then arrive at the result of row-reducing a column of symbols where the vector of constants usually resides. Since the row-reduced version of A has two zero rows, for a consistent system we require that

 $b_1 + 3b_3 - b_4 + 3b_5 + b_6 = 0$ $b_2 - 2b_3 + b_4 + b_5 - b_6 = 0$

Now we are exactly back where we were on the first go-round. Notice that we obtain the matrix L as simply the last two rows and last six columns of N.

This example motivates the remainder of this section, so it is worth careful study. You might attempt to mimic the second approach with the coefficient matrices of Archetype I [1714] and Archetype J [1723]. We will see shortly that the matrix L contains more information about A than just the column space.

Subsection EEF Extended echelon form

The final matrix that we row-reduced in Example CSANS [612] should look familiar in most respects to the procedure we used to compute the inverse of a nonsingular matrix, Theorem

CINM [516]. We will now generalize that procedure to matrices that are not necessarily nonsingular, or even square. First a definition.

Definition EEF Extended Echelon Form

Suppose A is an $m \times n$ matrix. Extend A on its right side with the addition of an $m \times m$ identity matrix to form an $m \times (n+m)$ matrix M. Use row operations to bring M to reduced row-echelon form and call the result N. N is the **extended reduced row-echelon form** of A, and we will standardize on names for five submatrices (B, C, J, K, L) of N.

Let B denote the $m \times n$ matrix formed from the first n columns of N and let J denote the $m \times m$ matrix formed from the last m columns of N. Suppose that B has r nonzero rows. Further partition N by letting C denote the $r \times n$ matrix formed from all of the non-zero rows of B. Let K be the $r \times m$ matrix formed from the first r rows of J, while L will be the $(m - r) \times m$ matrix formed from the bottom m - r rows of J. Pictorially,

$$M = [A|I_m] \xrightarrow{\text{RREF}} N = [B|J] = \begin{bmatrix} C & K \\ \hline 0 & L \end{bmatrix}$$

	٨	
/	1	١

Example SEEF Submatrices of extended echelon form We illustrate Definition EEF [619] with the matrix A,

$$A = \begin{bmatrix} 1 & -1 & -2 & 7 & 1 & 6 \\ -6 & 2 & -4 & -18 & -3 & -26 \\ 4 & -1 & 4 & 10 & 2 & 17 \\ 3 & -1 & 2 & 9 & 1 & 12 \end{bmatrix}$$

Augmenting with the 4×4 identity matrix, M=

$$\begin{bmatrix} 1 & -1 & -2 & 7 & 1 & 6 & 1 & 0 & 0 & 0 \\ -6 & 2 & -4 & -18 & -3 & -26 & 0 & 1 & 0 & 0 \\ 4 & -1 & 4 & 10 & 2 & 17 & 0 & 0 & 1 & 0 \\ 3 & -1 & 2 & 9 & 1 & 12 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and row-reducing, we obtain

$$N = \begin{bmatrix} 1 & 0 & 2 & 1 & 0 & 3 & 0 & 1 & 1 & 1 \\ 0 & 1 & 4 & -6 & 0 & -1 & 0 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 0 & -1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 2 & 1 \end{bmatrix}$$

So we then obtain

$$B = \begin{bmatrix} 1 & 0 & 2 & 1 & 0 & 3 \\ 0 & 1 & 4 & -6 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 2 & 1 & 0 & 3 \\ 0 & 1 & 4 & -6 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$
$$J = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 2 & 3 & 0 \\ 0 & -1 & 0 & -2 \\ 1 & 2 & 2 & 1 \end{bmatrix}$$
$$K = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 2 & 3 & 0 \\ 0 & -1 & 0 & -2 \end{bmatrix}$$
$$L = \begin{bmatrix} 1 & 2 & 2 & 1 \end{bmatrix}$$

You can observe (or verify) the properties of the following theorem with this example. $\quad \boxtimes$

Theorem PEEF Properties of Extended Echelon Form

Suppose that A is an $m \times n$ matrix and that N is its extended echelon form. Then

1. J is nonsingular.

2. B = JA.

3. If
$$\mathbf{x} \in \mathbb{C}^n$$
 and $\mathbf{y} \in \mathbb{C}^m$, then $A\mathbf{x} = \mathbf{y}$ if and only if $B\mathbf{x} = J\mathbf{y}$.

- 4. C is in reduced row-echelon form, has no zero rows and has r pivot columns.
- 5. L is in reduced row-echelon form, has no zero rows and has m r pivot columns.

Proof J is the result of applying a sequence of row operations to I_m , as such J and I_m are row-equivalent. $\mathcal{LS}(I_m, \mathbf{0})$ has only the zero solution, since I_m is nonsingular (Theorem NMRRI [176]). Thus, $\mathcal{LS}(J, \mathbf{0})$ also has only the zero solution (Theorem REMES [61], Definition ESYS [22]) and J is therefore nonsingular (Definition NSM [152]).

To prove the second part of this conclusion, first convince yourself that row operations and the matrix-vector are commutative operations. By this we mean the following. Suppose that F is an $m \times n$ matrix that is row-equivalent to the matrix G. Apply to the column vector $F\mathbf{w}$ the same sequence of row operations that converts F to G. Then the result is $G\mathbf{w}$. So we can do row operations on the matrix, then do a matrix-vector product, or do a matrix-vector product and then do row operations on a column vector, and the result will be the same either way. Since matrix multiplication is defined by a collection of matrix-vector products (Definition MM [468]), if we apply to the matrix product FH the same sequence of row operations that converts F to G then the result will equal GH. Now apply these observations to A.

Write $AI_n = I_m A$ and apply the row operations that convert M to N. A is converted to B, while I_m is converted to J, so we have $BI_n = JA$. Simplifying the left side gives the desired conclusion.

For the third conclusion, we now establish the two equivalences

$$A\mathbf{x} = \mathbf{y} \quad \iff \quad JA\mathbf{x} = J\mathbf{y} \quad \iff \quad B\mathbf{x} = J\mathbf{y}$$

The forward direction of the first equivalence is accomplished by multiplying both sides of the matrix equality by J, while the backward direction is accomplished by multiplying by the inverse of J (which we know exists by Theorem NI [542] since J is nonsingular). The second equivalence is obtained simply by the substitutions given by JA = B.

The first r rows of N are in reduced row-echelon form, since any contiguous collection of rows taken from a matrix in reduced row-echelon form will form a matrix that is again in reduced row-echelon form. Since the matrix C is formed by removing the last n entries of each these rows, the remainder is still in reduced row-echelon form. By its construction, Chas no zero rows. C has r rows and each contains a leading 1, so there are r pivot columns in C.

The final m-r rows of N are in reduced row-echelon form, since any contiguous collection of rows taken from a matrix in reduced row-echelon form will form a matrix that is again in reduced row-echelon form. Since the matrix L is formed by removing the first n entries of each these rows, and these entries are all zero (they form the zero rows of B), the remainder is still in reduced row-echelon form. L is the final m-r rows of the nonsingular matrix J, so none of these rows can be totally zero, or J would not row-reduce to the identity matrix. L has m-r rows and each contains a leading 1, so there are m-r pivot columns in L.

Notice that in the case where A is a nonsingular matrix we know that the reduced rowechelon form of A is the identity matrix (Theorem NMRRI [176]), so $B = I_n$. Then the second conclusion above says $JA = B = I_n$, so J is the inverse of A. Thus this theorem generalizes Theorem CINM [516], though the result is a "left-inverse" of A rather than a "right-inverse."

The third conclusion of Theorem PEEF [621] is the most telling. It says that \mathbf{x} is a solution to the linear system $\mathcal{LS}(A, \mathbf{y})$ if and only if \mathbf{x} is a solution to the linear system $\mathcal{LS}(B, J\mathbf{y})$. Or said differently, if we row-reduce the augmented matrix $[A \mid \mathbf{y}]$ we will get the augmented matrix $[B \mid J\mathbf{y}]$. The matrix J tracks the cumulative effect of the row operations that converts A to reduced row-echelon form, here effectively applying them to the vector of constants in a system of equations having A as a coefficient matrix. When A row-reduces to a matrix with zero rows, then $J\mathbf{y}$ should also have zero entries in the same rows if the system is to be consistent.

Subsection FS Four Subsets

With all the preliminaries in place we can state our main result for this section. In essence this result will allow us to say that we can find linearly independent sets to use in span constructions for all four subsets (null space, column space, row space, left null space) by analyzing only the extended echelon form of the matrix, and specifically, just the two submatrices C and L, which will be ripe for analysis since they are already in reduced row-echelon form (Theorem PEEF [621]).

Theorem FS Four Subsets

Suppose A is an $m \times n$ matrix with extended echelon form N. Suppose the reduced rowechelon form of A has r nonzero rows. Then C is the submatrix of N formed from the first r rows and the first n columns and L is the submatrix of N formed from the last m columns and the last m - r rows. Then

- 1. The null space of A is the null space of C, $\mathcal{N}(A) = \mathcal{N}(C)$.
- 2. The row space of A is the row space of C, $\mathcal{R}(A) = \mathcal{R}(C)$.
- 3. The column space of A is the null space of L, $C(A) = \mathcal{N}(L)$.
- 4. The left null space of A is the row space of L, $\mathcal{L}(A) = \mathcal{R}(L)$.

Proof First, $\mathcal{N}(A) = \mathcal{N}(B)$ since *B* is row-equivalent to *A* (Theorem REMES [61]). The zero rows of *B* represent equations that are always true in the homogeneous system $\mathcal{LS}(B, \mathbf{0})$, so the removal of these equations will not change the solution set. Thus, in turn, $\mathcal{N}(B) = \mathcal{N}(C)$.

Second, $\mathcal{R}(A) = \mathcal{R}(B)$ since *B* is row-equivalent to *A* (Theorem REMRS [580]). The zero rows of *B* contribute nothing to the span that is the row space of *B*, so the removal of these rows will not change the row space. Thus, in turn, $\mathcal{R}(B) = \mathcal{R}(C)$.

Third, we prove the set equality $\mathcal{C}(A) = \mathcal{N}(L)$ with Definition SE [1604]. Begin by showing that $\mathcal{C}(A) \subseteq \mathcal{N}(L)$. Choose $\mathbf{y} \in \mathcal{C}(A) \subseteq \mathbb{C}^m$. Then there exists a vector $\mathbf{x} \in \mathbb{C}^n$ such that $A\mathbf{x} = \mathbf{y}$ (Theorem CSCS [563]). Then for $1 \leq k \leq m - r$,

$[L\mathbf{y}]_k = [J\mathbf{y}]_{r+k}$	L a submatrix of J
$= [B\mathbf{x}]_{r+k}$	Theorem PEEF [621]
$=\left[\mathcal{O}\mathbf{x} ight]_k$	Zero matrix a submatrix of B
$= [0]_k$	Theorem MMZM $[474]$

So, for all $1 \leq k \leq m - r$, $[L\mathbf{y}]_k = [\mathbf{0}]_k$. So by Definition CVE [199] we have $L\mathbf{y} = \mathbf{0}$ and thus $\mathbf{y} \in \mathcal{N}(L)$.

Now, show that $\mathcal{N}(L) \subseteq \mathcal{C}(A)$. Choose $\mathbf{y} \in \mathcal{N}(L) \subseteq \mathbb{C}^m$. Form the vector $K\mathbf{y} \in \mathbb{C}^r$. The linear system $\mathcal{LS}(C, K\mathbf{y})$ is consistent since C is in reduced row-echelon form and has no zero rows (Theorem PEEF [621]). Let $\mathbf{x} \in \mathbb{C}^n$ denote a solution to $\mathcal{LS}(C, K\mathbf{y})$.

Then for $1 \leq j \leq r$,

$[B\mathbf{x}]_j = [C\mathbf{x}]_j$	${\cal C}$ a submatrix of ${\cal B}$
$= [K\mathbf{y}]_j$	\mathbf{x} a solution to $\mathcal{LS}(C, K\mathbf{y})$
$=\left[J\mathbf{y} ight]_{j}$	K a submatrix of J

And for $r + 1 \le k \le m$, $[B\mathbf{x}]_k = [\mathcal{O}\mathbf{x}]_{k-r}$ Zero matrix a submatrix of B

$$= [\mathbf{0}]_{k-r}$$

$$= [L\mathbf{y}]_{k-r}$$

$$= [J\mathbf{y}]_{k}$$
Theorem MMZM [474]
 $\mathbf{y} \text{ in } \mathcal{N}(L)$

$$L \text{ a submatrix of } J$$

So for all $1 \leq i \leq m$, $[B\mathbf{x}]_i = [J\mathbf{y}]_i$ and by Definition CVE [199] we have $B\mathbf{x} = J\mathbf{y}$. From Theorem PEEF [621] we know then that $A\mathbf{x} = \mathbf{y}$, and therefore $\mathbf{y} \in \mathcal{C}(A)$ (Theorem CSCS [563]). By Definition SE [1604] we now have $\mathcal{C}(A) = \mathcal{N}(L)$.

Fourth, we prove the set equality $\mathcal{L}(A) = \mathcal{R}(L)$ with Definition SE [1604]. Begin by showing that $\mathcal{R}(L) \subseteq \mathcal{L}(A)$. Choose $\mathbf{y} \in \mathcal{R}(L) \subseteq \mathbb{C}^m$. Then there exists a vector $\mathbf{w} \in \mathbb{C}^{m-r}$ such that $\mathbf{y} = L^t \mathbf{w}$ (Definition RSM [577], Theorem CSCS [563]). Then for $1 \leq i \leq n$,

m

$$\begin{bmatrix} A^{t} \mathbf{y} \end{bmatrix}_{i} = \sum_{k=1}^{m} \begin{bmatrix} A^{t} \end{bmatrix}_{ik} \begin{bmatrix} \mathbf{y} \end{bmatrix}_{k} \qquad \text{Theorem EMP [471]}$$
$$= \sum_{k=1}^{m} \begin{bmatrix} A^{t} \end{bmatrix}_{ik} \begin{bmatrix} L^{t} \mathbf{w} \end{bmatrix}_{k} \qquad \text{Definition of } \mathbf{w}$$
$$= \sum_{k=1}^{m} \begin{bmatrix} A^{t} \end{bmatrix}_{ik} \sum_{\ell=1}^{m-r} \begin{bmatrix} L^{t} \end{bmatrix}_{k\ell} \begin{bmatrix} \mathbf{w} \end{bmatrix}_{\ell} \qquad \text{Theorem EMP [471]}$$
$$= \sum_{k=1}^{m} \sum_{\ell=1}^{m-r} \begin{bmatrix} A^{t} \end{bmatrix}_{ik} \begin{bmatrix} L^{t} \end{bmatrix}_{k\ell} \begin{bmatrix} \mathbf{w} \end{bmatrix}_{\ell} \qquad \text{Property DCN [1597]}$$

$$= \sum_{\ell=1}^{m-r} \sum_{k=1}^{m} [A^{t}]_{ik} [L^{t}]_{k\ell} [\mathbf{w}]_{\ell} \qquad \text{Property CACN [1597]}$$

$$= \sum_{\ell=1}^{m-r} \left(\sum_{k=1}^{m} [A^{t}]_{ik} [L^{t}]_{k\ell} \right) [\mathbf{w}]_{\ell} \qquad \text{Property DCN [1597]}$$

$$= \sum_{\ell=1}^{m-r} \left(\sum_{k=1}^{m} [A^{t}]_{ik} [J^{t}]_{k,r+\ell} \right) [\mathbf{w}]_{\ell} \qquad L \text{ a submatrix of } J$$

$$= \sum_{\ell=1}^{m-r} [A^{t}J^{t}]_{i,r+\ell} [\mathbf{w}]_{\ell} \qquad \text{Theorem EMP [471]}$$

$$= \sum_{\ell=1}^{m-r} [(JA)^{t}]_{i,r+\ell} [\mathbf{w}]_{\ell} \qquad \text{Theorem MMT [482]}$$

$$= \sum_{\ell=1}^{m-r} [B^{t}]_{i,r+\ell} [\mathbf{w}]_{\ell} \qquad \text{Theorem PEEF [621]}$$

$$= \sum_{\ell=1}^{m-r} 0 [\mathbf{w}]_{\ell} \qquad \text{Zero rows in } B$$

$$= 0 \qquad \text{Property ZCN [1597]}$$

$$= [\mathbf{0}]_{i} \qquad \text{Definition ZCV [53]}$$

Since $[A^t \mathbf{y}]_i = [\mathbf{0}]_i$ for $1 \le i \le n$, Definition CVE [199] implies that $A^t \mathbf{y} = \mathbf{0}$. This means

that $\mathbf{y} \in \mathcal{N}(A^t)$.

Now, show that $\mathcal{L}(A) \subseteq \mathcal{R}(L)$. Choose $\mathbf{y} \in \mathcal{L}(A) \subseteq \mathbb{C}^m$. The matrix J is nonsingular (Theorem PEEF [621]), so J^t is also nonsingular (Theorem MIT [522]) and therefore the linear system $\mathcal{LS}(J^t, \mathbf{y})$ has a unique solution. Denote this solution as $\mathbf{x} \in \mathbb{C}^m$. We will need to work with two "halves" of \mathbf{x} , which we will denote as \mathbf{z} and \mathbf{w} with formal definitions given by

 $[z]_j = [x]_i \qquad 1 \le j \le r, \qquad [w]_k = [x]_{r+k} \qquad 1 \le k \le m-r$ Now, for $1 \le j \le r$,

$$\begin{bmatrix} C^{t} \mathbf{z} \end{bmatrix}_{j} = \sum_{k=1}^{r} \begin{bmatrix} C^{t} \end{bmatrix}_{jk} [\mathbf{z}]_{k} \qquad \text{Theorem EMP [471]}$$

$$= \sum_{k=1}^{r} \begin{bmatrix} C^{t} \end{bmatrix}_{jk} [\mathbf{z}]_{k} + \sum_{\ell=1}^{m-r} [\mathcal{O}]_{j\ell} [\mathbf{w}]_{\ell} \qquad \text{Definition ZM [433]}$$

$$= \sum_{k=1}^{r} \begin{bmatrix} B^{t} \end{bmatrix}_{jk} [\mathbf{z}]_{k} + \sum_{\ell=1}^{m-r} \begin{bmatrix} B^{t} \end{bmatrix}_{j,r+\ell} [\mathbf{w}]_{\ell} \qquad C, \mathcal{O} \text{ submatrices of } B$$

$$= \sum_{k=1}^{r} \begin{bmatrix} B^{t} \end{bmatrix}_{jk} [\mathbf{x}]_{k} + \sum_{\ell=1}^{m-r} \begin{bmatrix} B^{t} \end{bmatrix}_{j,r+\ell} [\mathbf{x}]_{r+\ell} \qquad \text{Definitions of } \mathbf{z} \text{ and } \mathbf{w}$$

$$= \sum_{k=1}^{r} \begin{bmatrix} B^{t} \end{bmatrix}_{jk} [\mathbf{x}]_{k} + \sum_{k=r+1}^{m} \begin{bmatrix} B^{t} \end{bmatrix}_{jk} [\mathbf{x}]_{k} \qquad \text{Re-index second sum}$$

$$= \sum_{k=1}^{m} [B^{t}]_{jk} [\mathbf{x}]_{k}$$

$$= \sum_{k=1}^{m} [(JA)^{t}]_{jk} [\mathbf{x}]_{k}$$

$$= \sum_{k=1}^{m} [A^{t}J^{t}]_{jk} [\mathbf{x}]_{k}$$

$$= \sum_{k=1}^{m} \sum_{\ell=1}^{m} [A^{t}]_{j\ell} [J^{t}]_{\ell k} [\mathbf{x}]_{k}$$

$$= \sum_{\ell=1}^{m} \sum_{k=1}^{m} [A^{t}]_{j\ell} [J^{t}]_{\ell k} [\mathbf{x}]_{k}$$

$$= \sum_{\ell=1}^{m} [A^{t}]_{j\ell} \left(\sum_{k=1}^{m} [J^{t}]_{\ell k} [\mathbf{x}]_{k}\right)$$

$$= \sum_{\ell=1}^{m} [A^{t}]_{j\ell} [J^{t}\mathbf{x}]_{\ell}$$

$$= \sum_{\ell=1}^{m} [A^{t}]_{j\ell} [J^{t}\mathbf{x}]_{\ell}$$

$$= [\mathbf{0}]_j \qquad \qquad \mathbf{y} \in \mathcal{L}(A$$

So, by Definition CVE [199], $C^t \mathbf{z} = \mathbf{0}$ and the vector \mathbf{z} gives us a linear combination of the columns of C^t that equals the zero vector. In other words, \mathbf{z} gives a relation of linear dependence on the rows of C. However, the rows of C are a linearly independent set by Theorem BRS [583]. According to Definition LICV [317] we must conclude that the entries of \mathbf{z} are all zero, i.e. $\mathbf{z} = \mathbf{0}$.

Now, for $1 \leq i \leq m$, we have

$$\begin{aligned} \left[\mathbf{y} \right]_{i} &= \left[J^{t} \mathbf{x} \right]_{i} & \text{Definition of } \mathbf{x} \\ &= \sum_{k=1}^{m} \left[J^{t} \right]_{ik} \left[\mathbf{x} \right]_{k} & \text{Theorem EMP [471]} \\ &= \sum_{k=1}^{r} \left[J^{t} \right]_{ik} \left[\mathbf{x} \right]_{k} + \sum_{k=r+1}^{m} \left[J^{t} \right]_{ik} \left[\mathbf{x} \right]_{k} & \text{Break apart sum} \\ &= \sum_{k=1}^{r} \left[J^{t} \right]_{ik} \left[\mathbf{z} \right]_{k} + \sum_{k=r+1}^{m} \left[J^{t} \right]_{ik} \left[\mathbf{w} \right]_{k-r} & \text{Definition of } \mathbf{z} \text{ and } \mathbf{w} \\ &= \sum_{k=1}^{r} \left[J^{t} \right]_{ik} 0 + \sum_{\ell=1}^{m-r} \left[J^{t} \right]_{i,r+\ell} \left[\mathbf{w} \right]_{\ell} & \mathbf{z} = \mathbf{0}, \text{ re-index} \\ &= 0 + \sum_{\ell=1}^{m-r} \left[L^{t} \right]_{i,\ell} \left[\mathbf{w} \right]_{\ell} & L \text{ a submatrix of } J \end{aligned}$$

$$= \left[L^{t} \mathbf{w} \right]_{i}$$
 Theorem EMP [471]

So by Definition CVE [199], $\mathbf{y} = L^t \mathbf{w}$. The existence of \mathbf{w} implies that $\mathbf{y} \in \mathcal{R}(L)$, and therefore $\mathcal{L}(A) \subseteq \mathcal{R}(L)$. So by Definition SE [1604] we have $\mathcal{L}(A) = \mathcal{R}(L)$.

The first two conclusions of this theorem are nearly trivial. But they set up a pattern of results for C that is reflected in the latter two conclusions about L. In total, they tell us that we can compute all four subsets just by finding null spaces and row spaces. This theorem does not tell us exactly how to compute these subsets, but instead simply expresses them as null spaces and row spaces of matrices in reduced row-echelon form without any zero rows (C and L). A linearly independent set that spans the null space of a matrix in reduced row-echelon form can be found easily with Theorem BNS [334]. It is an even easier matter to find a linearly independent set that spans the row space of a matrix in reduced row-echelon form with Theorem BRS [583], especially when there are no zero rows present. So an application of Theorem FS [625] is typically followed by two applications each of Theorem BNS [334].

The situation when r = m deserves comment, since now the matrix L has no rows. What is $\mathcal{C}(A)$ when we try to apply Theorem FS [625] and encounter $\mathcal{N}(L)$? One interpretation of this situation is that L is the coefficient matrix of a homogeneous system that has no equations. How hard is it to find a solution vector to this system? Some thought will convince you that *any* proposed vector will qualify as a solution, since it makes *all* of the equations true. So every possible vector is in the null space of L and therefore $\mathcal{C}(A) = \mathcal{N}(L) = \mathbb{C}^m$. OK, perhaps this sounds like some twisted argument from *Alice in Wonderland*. Let us try another argument that might solidly convince you of this logic.

If r = m, when we row-reduce the augmented matrix of $\mathcal{LS}(A, \mathbf{b})$ the result will have no zero rows, and all the leading 1's will occur in first *n* columns, so by Theorem RCLS [120] the system will be consistent. By Theorem CSCS [563], $\mathbf{b} \in \mathcal{C}(A)$. Since **b** was arbitrary, every possible vector is in the column space of *A*, so we again have $\mathcal{C}(A) = \mathbb{C}^m$. The situation when a matrix has r = m is known by the term **full rank**, and in the case of a square matrix coincides with nonsingularity (see Exercise FS.M50 [649]).

The properties of the matrix L described by this theorem can be explained informally as follows. A column vector $\mathbf{y} \in \mathbb{C}^m$ is in the column space of A if the linear system $\mathcal{LS}(A, \mathbf{y})$ is consistent (Theorem CSCS [563]). By Theorem RCLS [120], the reduced row-echelon form of the augmented matrix $[A \mid \mathbf{y}]$ of a consistent system will have zeros in the bottom m - rlocations of the last column. By Theorem PEEF [621] this final column is the vector $J\mathbf{y}$ and so should then have zeros in the final m - r locations. But since L comprises the final m - rrows of J, this condition is expressed by saying $\mathbf{y} \in \mathcal{N}(L)$.

Additionally, the rows of J are the scalars in linear combinations of the rows of A that create the rows of B. That is, the rows of J record the net effect of the sequence of row operations that takes A to its reduced row-echelon form, B. This can be seen in the equation JA = B (Theorem PEEF [621]). As such, the rows of L are scalars for linear combinations of the rows of A that yield zero rows. But such linear combinations are precisely the elements of the left null space. So any element of the row space of L is also an element of the left null space of A. We will now illustrate Theorem FS [625] with a few examples.

Example FS1 Four subsets, #1 In Example SEEF [619] we found the five relevant submatrices of the matrix

$$A = \begin{bmatrix} 1 & -1 & -2 & 7 & 1 & 6 \\ -6 & 2 & -4 & -18 & -3 & -26 \\ 4 & -1 & 4 & 10 & 2 & 17 \\ 3 & -1 & 2 & 9 & 1 & 12 \end{bmatrix}$$

To apply Theorem FS [625] we only need C and L,

$$C = \begin{bmatrix} 1 & 0 & 2 & 1 & 0 & 3 \\ 0 & 1 & 4 & -6 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix} \qquad \qquad L = \begin{bmatrix} 1 & 2 & 2 & 1 \end{bmatrix}$$

Then we use Theorem FS [625] to obtain

$$\mathcal{N}(A) = \mathcal{N}(C) = \left\langle \left\{ \begin{bmatrix} -2\\ -4\\ 1\\ 0\\ 0\\ 0 \end{bmatrix}, \begin{bmatrix} -1\\ 6\\ 0\\ 1\\ 0\\ 0 \end{bmatrix}, \begin{bmatrix} -3\\ 1\\ 0\\ 0\\ -2\\ 1 \end{bmatrix} \right\} \right\rangle$$

Theorem BNS [334]

$$\mathcal{R}(A) = \mathcal{R}(C) = \left\langle \left\{ \begin{bmatrix} 1\\0\\2\\1\\0\\0\\3 \end{bmatrix}, \begin{bmatrix} 0\\1\\4\\-6\\0\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\0\\1\\2 \end{bmatrix} \right\} \right\rangle$$
 Theorem BRS [583]
$$\mathcal{C}(A) = \mathcal{N}(L) = \left\langle \left\{ \begin{bmatrix} -2\\1\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} -2\\0\\1\\0\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} -1\\0\\0\\1\\0\\1 \end{bmatrix} \right\} \right\rangle$$
 Theorem BNS [334]
$$\mathcal{L}(A) = \mathcal{R}(L) = \left\langle \left\{ \begin{bmatrix} 1\\2\\2\\1\\1\\2\\1\\1 \end{bmatrix} \right\} \right\rangle$$

Boom!

 \boxtimes

Example FS2 Four subsets, #2

Now lets return to the matrix A that we used to motivate this section in Example CSANS

Version 2.99

[612],

$$A = \begin{bmatrix} 10 & 0 & 3 & 8 & 7 \\ -16 & -1 & -4 & -10 & -13 \\ -6 & 1 & -3 & -6 & -6 \\ 0 & 2 & -2 & -3 & -2 \\ 3 & 0 & 1 & 2 & 3 \\ -1 & -1 & 1 & 1 & 0 \end{bmatrix}$$

We form the matrix M by adjoining the 6×6 identity matrix I_6 ,

$$M = \begin{bmatrix} 10 & 0 & 3 & 8 & 7 & 1 & 0 & 0 & 0 & 0 \\ -16 & -1 & -4 & -10 & -13 & 0 & 1 & 0 & 0 & 0 & 0 \\ -6 & 1 & -3 & -6 & -6 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & -2 & -3 & -2 & 0 & 0 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 2 & 3 & 0 & 0 & 0 & 0 & 1 & 0 \\ -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and row-reduce to obtain ${\cal N}$

$$N = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 & 0 & 0 & 1 & -1 & 2 & -1 \\ 0 & 1 & 0 & 0 & -3 & 0 & 0 & -2 & 3 & -3 & 3 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 3 & 3 \\ 0 & 0 & 0 & 1 & -2 & 0 & 0 & -2 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 3 & -1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 & 1 & -1 \end{bmatrix}$$

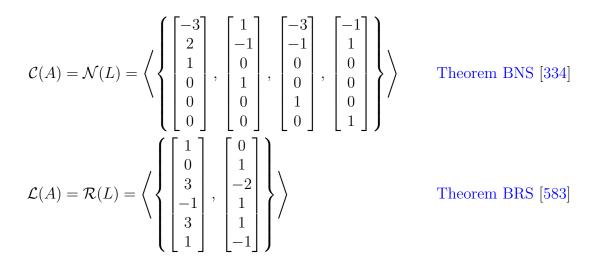
To find the four subsets for A, we only need identify the 4×5 matrix C and the 2×6 matrix L,

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & -3 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & 0 & 3 & -1 & 3 & 1 \\ 0 & 1 & -2 & 1 & 1 & -1 \end{bmatrix}$$

Then we apply Theorem FS [625],

$$\mathcal{N}(A) = \mathcal{N}(C) = \left\langle \left\{ \begin{bmatrix} -2\\3\\-1\\2\\1 \end{bmatrix} \right\} \right\rangle$$
 Theorem BNS [334]
$$\mathcal{R}(A) = \mathcal{R}(C) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0\\2\\2 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0\\-3\\2 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\1\\2\\3 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\1\\2\\3 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\1\\2\\3 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\1\\2\\2\\3 \end{bmatrix} \right\} \right\rangle$$
 Theorem BRS [583]



 \boxtimes

The next example is just a bit different since the matrix has more rows than columns, and a trivial null space.

Example FSAG Four subsets, Archetype G

Archetype G [1697] and Archetype H [1705] are both systems of m = 5 equations in n = 2 variables. They have identical coefficient matrices, which we will denote here as the matrix

G,

$$G = \begin{bmatrix} 2 & 3 \\ -1 & 4 \\ 3 & 10 \\ 3 & -1 \\ 6 & 9 \end{bmatrix}$$

Adjoin the 5×5 identity matrix, I_5 , to form

$$M = \begin{bmatrix} 2 & 3 & 1 & 0 & 0 & 0 & 0 \\ -1 & 4 & 0 & 1 & 0 & 0 & 0 \\ 3 & 10 & 0 & 0 & 1 & 0 & 0 \\ 3 & -1 & 0 & 0 & 0 & 1 & 0 \\ 6 & 9 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This row-reduces to

$$N = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \frac{3}{11} & \frac{1}{33} \\ 0 & 1 & 0 & 0 & 0 & -\frac{2}{11} & \frac{1}{11} \\ 0 & 0 & 1 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 1 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

The first n = 2 columns contain r = 2 leading 1's, so we obtain C as the 2×2 identity

matrix and extract L from the final m - r = 3 rows in the final m = 5 columns.

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \qquad L = \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

Then we apply Theorem FS [625],

$$\mathcal{N}(G) = \mathcal{N}(C) = \langle \emptyset \rangle = \{\mathbf{0}\}$$
$$\mathcal{R}(G) = \mathcal{R}(C) = \left\langle \left\{ \begin{bmatrix} 1\\0 \end{bmatrix}, \begin{bmatrix} 0\\1 \end{bmatrix} \right\} \right\rangle = \mathbb{C}^2$$
$$\mathcal{C}(G) = \mathcal{N}(L) = \left\langle \left\{ \begin{bmatrix} 0\\-1\\-1\\1\\0 \end{bmatrix}, \begin{bmatrix} \frac{1}{3}\\\frac{1}{3}\\1\\0\\1 \end{bmatrix} \right\} \right\rangle$$
$$= \left\langle \left\{ \begin{bmatrix} 0\\-1\\-1\\1\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\3\\0\\3 \end{bmatrix} \right\} \right\rangle$$

Theorem BNS [334]

Theorem BRS [583]

Theorem BNS [334]

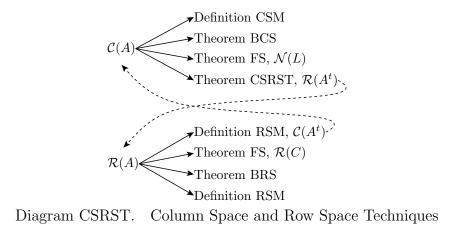
Theorem BRS [583]

As mentioned earlier, Archetype G [1697] is consistent, while Archetype H [1705] is inconsistent. See if you can write the two different vectors of constants from these two archetypes as linear combinations of the two vectors in $\mathcal{C}(G)$. How about the two columns of G, can you write each individually as a linear combination of the two vectors in $\mathcal{C}(G)$? They must be in the column space of G also. Are your answers unique? Do you notice anything about the scalars that appear in the linear combinations you are forming?

Example COV [370] and Example CSROI [587] each describes the column space of the coefficient matrix from Archetype I [1714] as the span of a set of r = 3 linearly independent vectors. It is no accident that these two different sets both have the same size. If we (you?) were to calculate the column space of this matrix using the null space of the matrix L from Theorem FS [625] then we would again find a set of 3 linearly independent vectors that span the range. More on this later.

So we have three different methods to obtain a description of the column space of a matrix as the span of a linearly independent set. Theorem BCS [569] is sometimes useful since the vectors it specifies are equal to actual columns of the matrix. Theorem BRS [583] and Theorem CSRST [586] combine to create vectors with lots of zeros, and strategically placed 1's near the top of the vector. Theorem FS [625] and the matrix L from the extended echelon form gives us a third method, which tends to create vectors with lots of zeros, and strategically placed 1's near the bottom of the vector. If we don't care about linear independence we can also appeal to Definition CSM [560] and simply express the column space as the span of all the columns of the matrix, giving us a fourth description.

With Theorem CSRST [586] and Definition RSM [577], we can compute column spaces with theorems about row spaces, and we can compute row spaces with theorems about row spaces, but in each case we must transpose the matrix first. At this point you may be overwhelmed by all the possibilities for computing column and row spaces. Diagram CSRST [642] is meant to help. For both the column space and row space, it suggests four techniques. One is to appeal to the definition, another yields a span of a linearly independent set, and a third uses Theorem FS [625]. A fourth suggests transposing the matrix and the dashed line implies that then the companion set of techniques can be applied. This can lead to a bit of silliness, since if you were to follow the dashed lines *twice* you would transpose the matrix twice, and by Theorem TT [437] would accomplish nothing productive.



Although we have many ways to describe a column space, notice that one tempting strategy will usually fail. It is not possible to simply row-reduce a matrix directly and then use the columns of the row-reduced matrix as a set whose span equals the column space. In other words, row operations *do not* preserve column spaces (however row operations do preserve row spaces, Theorem REMRS [580]). See Exercise CRS.M21 [599].

Subsection READ Reading Questions

1. Find a nontrivial element of the left null space of A.

$$A = \begin{bmatrix} 2 & 1 & -3 & 4 \\ -1 & -1 & 2 & -1 \\ 0 & -1 & 1 & 2 \end{bmatrix}$$

2. Find the matrices C and L in the extended echelon form of A.

$$A = \begin{bmatrix} -9 & 5 & -3\\ 2 & -1 & 1\\ -5 & 3 & -1 \end{bmatrix}$$

3. Why is Theorem FS [625] a great conclusion to Chapter M [425]?

Subsection EXC Exercises

C20 Example FSAG [638] concludes with several questions. Perform the analysis suggested by these questions.

Contributed by Robert Beezer

C25 Given the matrix A below, use the extended echelon form of A to answer each part of this problem. In each part, find a linearly independent set of vectors, S, so that the span of S, $\langle S \rangle$, equals the specified set of vectors.

$$A = \begin{bmatrix} -5 & 3 & -1 \\ -1 & 1 & 1 \\ -8 & 5 & -1 \\ 3 & -2 & 0 \end{bmatrix}$$

- (a) The row space of A, $\mathcal{R}(A)$.
- (b) The column space of A, $\mathcal{C}(A)$.
- (c) The null space of A, $\mathcal{N}(A)$.
- (d) The left null space of A, $\mathcal{L}(A)$.

Contributed by Robert Beezer Solution [650]

C26 For the matrix D below use the extended echelon form to find

- (a) a linearly independent set whose span is the column space of D.
- (b) a linearly independent set whose span is the left null space of D.

$$D = \begin{bmatrix} -7 & -11 & -19 & -15\\ 6 & 10 & 18 & 14\\ 3 & 5 & 9 & 7\\ -1 & -2 & -4 & -3 \end{bmatrix}$$

Contributed by Robert Beezer Solution [652]

C41 The following archetypes are systems of equations. For each system, write the vector of constants as a linear combination of the vectors in the span construction for the column space provided by Theorem FS [625] and Theorem BNS [334] (these vectors are listed for each of these archetypes).

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670] Archetype E [1678] Archetype F [1686] Archetype G [1697] Archetype H [1705] Archetype I [1714] Archetype J [1723]

Contributed by Robert Beezer

C43 The following archetypes are either matrices or systems of equations with coefficient matrices. For each matrix, compute the extended echelon form N and identify the matrices C and L. Using Theorem FS [625], Theorem BNS [334] and Theorem BRS [583] express the null space, the row space, the column space and left null space of each coefficient matrix as a span of a linearly independent set.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678] Archetype F [1686] Archetype G [1697]/Archetype H [1705] Archetype I [1714] Archetype J [1723] Archetype K [1734] Archetype L [1742]

Contributed by Robert Beezer

C60 For the matrix *B* below, find sets of vectors whose span equals the column space of $B(\mathcal{C}(B))$ and which individually meet the following extra requirements.

(a) The set illustrates the definition of the column space.

(b) The set is linearly independent and the members of the set are columns of B.

(c) The set is linearly independent with a "nice pattern of zeros and ones" at the *top* of each vector.

(d) The set is linearly independent with a "nice pattern of zeros and ones" at the bottom of each vector.

$$B = \begin{bmatrix} 2 & 3 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ -1 & 2 & 3 & -4 \end{bmatrix}$$

Contributed by Robert Beezer Solution [653]

C61 Let A be the matrix below, and find the indicated sets with the requested properties.

$$A = \begin{bmatrix} 2 & -1 & 5 & -3 \\ -5 & 3 & -12 & 7 \\ 1 & 1 & 4 & -3 \end{bmatrix}$$

(a) A linearly independent set S so that $\mathcal{C}(A) = \langle S \rangle$ and S is composed of columns of A.

(b) A linearly independent set S so that $C(A) = \langle S \rangle$ and the vectors in S have a nice pattern of zeros and ones at the top of the vectors.

(c) A linearly independent set S so that $C(A) = \langle S \rangle$ and the vectors in S have a nice pattern of zeros and ones at the bottom of the vectors.

(d) A linearly independent set S so that $\mathcal{R}(A) = \langle S \rangle$. Contributed by Robert Beezer Solution [655] **M50** Suppose that A is a nonsingular matrix. Extend the four conclusions of Theorem FS [625] in this special case and discuss connections with previous results (such as Theorem NME4 [575]).

Contributed by Robert Beezer

M51 Suppose that A is a singular matrix. Extend the four conclusions of Theorem FS [625] in this special case and discuss connections with previous results (such as Theorem NME4 [575]).

Contributed by Robert Beezer

Subsection SOL Solutions

C25 Contributed by Robert Beezer Statement [645] Add a 4×4 identity matrix to the right of A to form the matrix M and then row-reduce to the matrix N,

$$M = \begin{bmatrix} -5 & 3 & -1 & 1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ -8 & 5 & -1 & 0 & 0 & 1 & 0 \\ 3 & -2 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2 & 0 & 0 & -2 & -5 \\ 0 & 1 & 3 & 0 & 0 & -3 & -8 \\ 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 3 \end{bmatrix} = N$$

To apply Theorem FS [625] in each of these four parts, we need the two matrices,

$$C = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} \qquad \qquad L = \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & 3 \end{bmatrix}$$

(a)

$$\mathcal{R}(A) = \mathcal{R}(C)$$
$$= \left\langle \begin{bmatrix} 1\\0\\2 \end{bmatrix}, \begin{bmatrix} 0\\1\\3 \end{bmatrix} \right\rangle$$

Theorem FS [625]

Theorem BRS [583]

(b)

$$\mathcal{C}(A) = \mathcal{N}(L)$$
$$= \left\langle \begin{bmatrix} 1\\-1\\1\\0 \end{bmatrix}, \begin{bmatrix} 1\\-3\\0\\1 \end{bmatrix} \right\rangle$$

Theorem FS [625]

Theorem BNS [334]

(c)

$$\mathcal{N}(A) = \mathcal{N}(C)$$
$$= \left\langle \begin{bmatrix} -2\\ -3\\ 1 \end{bmatrix} \right\rangle$$

Theorem FS [625]

Theorem BNS [334]

(d)

$$\mathcal{L}(A) = \mathcal{R}(L)$$
$$= \left\langle \begin{bmatrix} 1\\0\\-1\\-1 \end{bmatrix}, \begin{bmatrix} 0\\1\\1\\3 \end{bmatrix} \right\rangle$$

Theorem FS [625]

Theorem BRS [583]

C26 Contributed by Robert Beezer Statement [646] For both parts, we need the extended echelon form of the matrix.

$$\begin{bmatrix} -7 & -11 & -19 & -15 & 1 & 0 & 0 & 0 \\ 6 & 10 & 18 & 14 & 0 & 1 & 0 & 0 \\ 3 & 5 & 9 & 7 & 0 & 0 & 1 & 0 \\ -1 & -2 & -4 & -3 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -2 & -1 & 0 & 0 & 2 & 5 \\ 0 & 1 & 3 & 2 & 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 \end{bmatrix}$$

From this matrix we extract the last two rows, in the last four columns to form the matrix L,

$$L = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 0 \end{bmatrix}$$

(a) By Theorem FS [625] and Theorem BNS [334] we have

$$\mathcal{C}(D) = \mathcal{N}(L) = \left\langle \left\{ \begin{bmatrix} -3\\2\\1\\0 \end{bmatrix}, \begin{bmatrix} -2\\0\\0\\1 \end{bmatrix} \right\} \right\rangle$$

(b) By Theorem FS [625] and Theorem BRS [583] we have

$$\mathcal{L}(D) = \mathcal{R}(L) = \left\langle \left\{ \begin{bmatrix} 1\\0\\3\\2 \end{bmatrix}, \begin{bmatrix} 0\\1\\-2\\0 \end{bmatrix} \right\} \right\rangle$$

C60 Contributed by Robert Beezer Statement [647]

(a) The definition of the column space is the span of the set of columns (Definition CSM [560]). So the desired set is just the four columns of B,

$$S = \left\{ \begin{bmatrix} 2\\1\\-1 \end{bmatrix}, \begin{bmatrix} 3\\1\\2 \end{bmatrix}, \begin{bmatrix} 1\\0\\3 \end{bmatrix}, \begin{bmatrix} 1\\1\\-4 \end{bmatrix} \right\}$$

(b) Theorem BCS [569] suggests row-reducing the matrix and using the columns of B that correspond to the pivot columns.

$$B \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So the pivot columns are numbered by elements of $D = \{1, 2\}$, so the requested set is

$$S = \left\{ \begin{bmatrix} 2\\1\\-1 \end{bmatrix}, \begin{bmatrix} 3\\1\\2 \end{bmatrix} \right\}$$

(c) We can find this set by row-reducing the transpose of B, deleting the zero rows, and using the nonzero rows as column vectors in the set. This is an application of Theorem

CSRST [586] followed by Theorem BRS [583].

$$B^{t} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

So the requested set is

$$S = \left\{ \begin{bmatrix} 1\\0\\3 \end{bmatrix}, \begin{bmatrix} 0\\1\\-7 \end{bmatrix} \right\}$$

(d) With the column space expressed as a null space, the vectors obtained via Theorem BNS [334] will be of the desired shape. So we first proceed with Theorem FS [625] and create the extended echelon form,

$$\begin{bmatrix} B \mid I_3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ -1 & 2 & 3 & -4 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 2 & 0 & \frac{2}{3} & \frac{-1}{3} \\ 0 & 1 & 1 & -1 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{RREF}}$$

So, employing Theorem FS [625], we have $\mathcal{C}(B) = \mathcal{N}(L)$, where

$$L = \begin{bmatrix} 1 & \frac{-7}{3} & \frac{-1}{3} \end{bmatrix}$$

We can find the desired set of vectors from Theorem BNS [334] as

$$S = \left\{ \begin{bmatrix} \frac{7}{3} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{3} \\ 0 \\ 1 \end{bmatrix} \right\}$$

C61 Contributed by Robert Beezer Statement [648]

(a) First find a matrix B that is row-equivalent to A and in reduced row-echelon form

$$B = \begin{bmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem BCS [569] we can choose the columns of A that correspond to dependent variables $(D = \{1, 2\})$ as the elements of S and obtain the desired properties. So

$$S = \left\{ \begin{bmatrix} 2\\-5\\1 \end{bmatrix}, \begin{bmatrix} -1\\3\\1 \end{bmatrix} \right\}$$

(b) We can write the column space of A as the row space of the transpose (Theorem CSRST [586]). So we row-reduce the transpose of A to obtain the row-equivalent matrix C in reduced row-echelon form

$$C = \begin{bmatrix} 1 & 0 & 8 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The nonzero rows (written as columns) will be a linearly independent set that spans the row space of A^t , by Theorem BRS [583], and the zeros and ones will be at the top of the vectors,

$$S = \left\{ \begin{bmatrix} 1\\0\\8 \end{bmatrix}, \begin{bmatrix} 0\\1\\3 \end{bmatrix} \right\}$$

(c) In preparation for Theorem FS [625], augment A with the 3×3 identity matrix I_3 and row-reduce to obtain the extended echelon form,

$$\begin{bmatrix} 1 & 0 & 3 & -2 & 0 & -\frac{1}{8} & \frac{3}{8} \\ 0 & 1 & 1 & -1 & 0 & \frac{1}{8} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & 1 & \frac{3}{8} & -\frac{1}{8} \end{bmatrix}$$

Then since the first four columns of row 3 are all zeros, we extract

$$L = \begin{bmatrix} 1 & \frac{3}{8} & -\frac{1}{8} \end{bmatrix}$$

Theorem FS [625] says that $\mathcal{C}(A) = \mathcal{N}(L)$. We can then use Theorem BNS [334] to construct the desired set S, based on the free variables with indices in $F = \{2, 3\}$ for the homogeneous system $\mathcal{LS}(L, \mathbf{0})$, so

$$S = \left\{ \begin{bmatrix} -\frac{3}{8} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{8} \\ 0 \\ 1 \end{bmatrix} \right\}$$

Notice that the zeros and ones are at the bottom of the vectors.

(d) This is a straightforward application of Theorem BRS [583]. Use the row-reduced matrix B from part (a), grab the nonzero rows, and write them as column vectors,

$$S = \left\{ \begin{bmatrix} 1\\0\\3\\-2 \end{bmatrix}, \begin{bmatrix} 0\\1\\1\\-1 \end{bmatrix} \right\}$$

Chapter VS Vector Spaces

We now have a computational toolkit in place and so we can begin our study of linear algebra in a more theoretical style.

Linear algebra is the study of two fundamental objects, vector spaces and linear transformations (see Chapter LT [1071]). This chapter will focus on the former. The power of mathematics is often derived from generalizing many different situations into one abstract formulation, and that is exactly what we will be doing throughout this chapter.

Section VS Vector Spaces

In this section we present a formal definition of a vector space, which will lead to an extra increment of abstraction. Once defined, we study its most basic properties.

Subsection VS Vector Spaces

Here is one of the two most important definitions in the entire course.

Definition VS Vector Space

Suppose that V is a set upon which we have defined two operations: (1) vector addition, which combines two elements of V and is denoted by "+", and (2) scalar multiplication, which combines a complex number with an element of V and is denoted by juxtaposition. Then V, along with the two operations, is a vector space over \mathbb{C} if the following ten properties hold.

- AC Additive Closure If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u} + \mathbf{v} \in V$.
- SC Scalar Closure If $\alpha \in \mathbb{C}$ and $\mathbf{u} \in V$, then $\alpha \mathbf{u} \in V$.
- C Commutativity If $\mathbf{u}, \mathbf{v} \in V$, then $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
- AA Additive Associativity If $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, then $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
- Z Zero Vector There is a vector, 0, called the zero vector, such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for all $\mathbf{u} \in V$.
- AI Additive Inverses If $\mathbf{u} \in V$, then there exists a vector $-\mathbf{u} \in V$ so that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
- SMA Scalar Multiplication Associativity If $\alpha, \beta \in \mathbb{C}$ and $\mathbf{u} \in V$, then $\alpha(\beta \mathbf{u}) = (\alpha \beta)\mathbf{u}$.
- DVA Distributivity across Vector Addition If $\alpha \in \mathbb{C}$ and $\mathbf{u}, \mathbf{v} \in V$, then $\alpha(\mathbf{u} + \mathbf{v}) = \alpha \mathbf{u} + \alpha \mathbf{v}$.
- DSA Distributivity across Scalar Addition If $\alpha, \beta \in \mathbb{C}$ and $\mathbf{u} \in V$, then $(\alpha + \beta)\mathbf{u} = \alpha \mathbf{u} + \beta \mathbf{u}$.

• O One If $\mathbf{u} \in V$, then $1\mathbf{u} = \mathbf{u}$.

The objects in V are called **vectors**, no matter what else they might really be, simply by virtue of being elements of a vector space. \triangle

Now, there are several important observations to make. Many of these will be easier to understand on a second or third reading, and especially after carefully studying the examples in Subsection VS.EVS [663].

An **axiom** is often a "self-evident" truth. Something so fundamental that we all agree it is true and accept it without proof. Typically, it would be the logical underpinning that we would begin to build theorems upon. Some might refer to the ten properties of Definition VS [659] as axioms, implying that a vector space is a very natural object and the ten properties are the essence of a vector space. We will instead emphasize that we will begin with a definition of a vector space. After studying the remainder of this chapter, you might return here and remind yourself how all our forthcoming theorems and definitions rest on this foundation.

As we will see shortly, the objects in V can be *anything*, even though we will call them vectors. We have been working with vectors frequently, but we should stress here that these have so far just been *column* vectors — scalars arranged in a columnar list of fixed length. In a similar vein, you have used the symbol "+" for many years to represent the addition of numbers (scalars). We have extended its use to the addition of column vectors and to the addition of matrices, and now we are going to recycle it even further and let it denote vector addition in *any* possible vector space. So when describing a new vector space, we will have

to *define* exactly what "+" is. Similar comments apply to scalar multiplication. Conversely, we can *define* our operations any way we like, so long as the ten properties are fulfilled (see Example CVS [670]).

In Definition VS [659], the scalars do not have to be complex numbers. They can come from what are called in more advanced mathematics, "fields" (see Section F [1829] for more on these objects). Examples of fields are the set of complex numbers, the set of real numbers, the set of rational numbers, and even the finite set of "binary numbers", $\{0, 1\}$. There are many, many others. In this case we would call V a vector space over (the field) F.

A vector space is composed of three objects, a set and two operations. Some would explicitly state in the definition that V must be a non-empty set, be we can infer this from Property Z [660], since the set cannot be empty and contain a vector that behaves as the zero vector. Also, we usually use the same symbol for both the set and the vector space itself. Do not let this convenience fool you into thinking the operations are secondary!

This discussion has either convinced you that we are really embarking on a new level of abstraction, or they have seemed cryptic, mysterious or nonsensical. You might want to return to this section in a few days and give it another read then. In any case, let's look at some concrete examples now.

Subsection EVS Examples of Vector Spaces

Our aim in this subsection is to give you a storehouse of examples to work with, to become comfortable with the ten vector space properties and to convince you that the multitude of examples justifies (at least initially) making such a broad definition as Definition VS [659]. Some of our claims will be justified by reference to previous theorems, we will prove some facts from scratch, and we will do one non-trivial example completely. In other places, our usual thoroughness will be neglected, so grab paper and pencil and play along.

Example VSCV

The vector space \mathbb{C}^m

Set: \mathbb{C}^m , all column vectors of size m, Definition VSCV [198]. Equality: Entry-wise, Definition CVE [199]. Vector Addition: The "usual" addition, given in Definition CVA [201]. Scalar Multiplication: The "usual" scalar multiplication, given in Definition CVSM [203].

Does this set with these operations fulfill the ten properties? Yes. And by design all we need to do is quote Theorem VSPCV [205]. That was easy. \boxtimes

Example VSM

The vector space of matrices, M_{mn}

Set: M_{mn} , the set of all matrices of size $m \times n$ and entries from \mathbb{C} , Example VSM [663].

Equality: Entry-wise, Definition ME [427].

Vector Addition: The "usual" addition, given in Definition MA [427].

Scalar Multiplication: The "usual" scalar multiplication, given in Definition MSM [428].

Does this set with these operations fulfill the ten properties? Yes. And all we need to do is quote Theorem VSPM [430]. Another easy one (by design). \boxtimes

So, the set of all matrices of a fixed size forms a vector space. That entitles us to call a matrix a vector, since a matrix is an element of a vector space. For example, if $A, B \in M_{3,4}$ then we call A and B "vectors," and we even use our previous notation for column vectors to refer to A and B. So we could legitimately write expressions like

$$\mathbf{u} + \mathbf{v} = A + B = B + A = \mathbf{v} + \mathbf{u}$$

This could lead to some confusion, but it is not too great a danger. But it is worth comment.

The previous two examples may be less than satisfying. We made all the relevant definitions long ago. And the required verifications were all handled by quoting old theorems. However, it is important to consider these two examples first. We have been studying vectors and matrices carefully (Chapter V [197], Chapter M [425]), and both objects, along with their operations, have certain properties in common, as you may have noticed in comparing Theorem VSPCV [205] with Theorem VSPM [430]. Indeed, it is these two theorems that motivate us to formulate the abstract definition of a vector space, Definition VS [659]. Now, should we prove some general theorems about vector spaces (as we will shortly in Subsection VS.VSP [674]), we can instantly apply the conclusions to both \mathbb{C}^m and M_{mn} . Notice too how we have taken six definitions and two theorems and reduced them down to two *examples*. With greater generalization and abstraction our old ideas get downgraded in stature.

Let us look at some more examples, now considering some new vector spaces.

Example VSP The vector space of polynov

The vector space of polynomials, P_n

Set: P_n , the set of all polynomials of degree n or less in the variable x with coefficients from \mathbb{C} .

Equality:

$$a_0 + a_1x + a_2x^2 + \dots + a_nx^n = b_0 + b_1x + b_2x^2 + \dots + b_nx^n$$
 if and only if $a_i = b_i$ for $0 \le i \le n$

Vector Addition:

$$(a_0 + a_1x + a_2x^2 + \dots + a_nx^n) + (b_0 + b_1x + b_2x^2 + \dots + b_nx^n) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \dots + (a_n + b_n)x^n$$

Scalar Multiplication:

$$\alpha(a_0 + a_1x + a_2x^2 + \dots + a_nx^n) = (\alpha a_0) + (\alpha a_1)x + (\alpha a_2)x^2 + \dots + (\alpha a_n)x^n$$

This set, with these operations, will fulfill the ten properties, though we will not work all the details here. However, we will make a few comments and prove one of the properties. First, the zero vector (Property Z [660]) is what you might expect, and you can check that it has the required property.

$$\mathbf{0} = 0 + 0x + 0x^2 + \dots + 0x^n$$

The additive inverse (Property AI [660]) is also no surprise, though consider how we have chosen to write it.

$$-(a_0 + a_1x + a_2x^2 + \dots + a_nx^n) = (-a_0) + (-a_1)x + (-a_2)x^2 + \dots + (-a_n)x^n$$

Now let's prove the associativity of vector addition (Property AA [660]). This is a bit tedious, though necessary. Throughout, the plus sign ("+") does triple-duty. You might ask yourself what each plus sign represents as you work through this proof.

$$\begin{aligned} \mathbf{u} + (\mathbf{v} + \mathbf{w}) \\ &= (a_0 + a_1 x + \dots + a_n x^n) + ((b_0 + b_1 x + \dots + b_n x^n) + (c_0 + c_1 x + \dots + c_n x^n)) \\ &= (a_0 + a_1 x + \dots + a_n x^n) + ((b_0 + c_0) + (b_1 + c_1) x + \dots + (b_n + c_n) x^n) \\ &= (a_0 + (b_0 + c_0)) + (a_1 + (b_1 + c_1)) x + \dots + (a_n + (b_n + c_n)) x^n \\ &= ((a_0 + b_0) + c_0) + ((a_1 + b_1) + c_1) x + \dots + ((a_n + b_n) + c_n) x^n \\ &= ((a_0 + b_0) + (a_1 + b_1) x + \dots + (a_n + b_n) x^n) + (c_0 + c_1 x + \dots + c_n x^n) \\ &= ((a_0 + a_1 x + \dots + a_n x^n) + (b_0 + b_1 x + \dots + b_n x^n)) + (c_0 + c_1 x + \dots + c_n x^n) \\ &= (\mathbf{u} + \mathbf{v}) + \mathbf{w} \end{aligned}$$

Notice how it is the application of the associativity of the (old) addition of complex numbers in the middle of this chain of equalities that makes the whole proof happen. The remainder is successive applications of our (new) definition of vector (polynomial) addition. Proving the remainder of the ten properties is similar in style and tedium. You might try proving the commutativity of vector addition (Property C [660]), or one of the distributivity properties (Property DVA [660], Property DSA [660]).

Example VSIS The vector space of infinite sequences Set: $\mathbb{C}^{\infty} = \{ (c_0, c_1, c_2, c_3, \ldots) | c_i \in \mathbb{C}, i \in \mathbb{N} \}.$ Equality:

$$(c_0, c_1, c_2, \ldots) = (d_0, d_1, d_2, \ldots)$$
 if and only if $c_i = d_i$ for all $i \ge 0$

Vector Addition:

$$(c_0, c_1, c_2, \ldots) + (d_0, d_1, d_2, \ldots) = (c_0 + d_0, c_1 + d_1, c_2 + d_2, \ldots)$$

Scalar Multiplication:

$$\alpha(c_0, c_1, c_2, c_3, \ldots) = (\alpha c_0, \alpha c_1, \alpha c_2, \alpha c_3, \ldots)$$

This should remind you of the vector space \mathbb{C}^m , though now our lists of scalars are written horizontally with commas as delimiters and they are allowed to be infinite in length. What does the zero vector look like (Property Z [660])? Additive inverses (Property AI [660])? Can you prove the associativity of vector addition (Property AA [660])? \boxtimes

Example VSF

The vector space of functions

Let X be any set. Set: $F = \{f | f : X \to \mathbb{C}\}.$ Equality: f = g if and only if f(x) = g(x) for all $x \in X$. Vector Addition: f + g is the function with outputs defined by (f + g)(x) = f(x) + g(x). Scalar Multiplication: αf is the function with outputs defined by $(\alpha f)(x) = \alpha f(x)$.

So this is the set of all functions of one variable that take elements of the set X to a complex number. You might have studied functions of one variable that take a real number to a real number, and that might be a more natural set to use as X. But since we are allowing our scalars to be complex numbers, we need to specify that the range of our functions is the complex numbers. Study carefully how the definitions of the operation are made, and think about the different uses of "+" and juxtaposition. As an example of what is required when verifying that this is a vector space, consider that the zero vector (Property Z [660]) is the function z whose definition is z(x) = 0 for every input $x \in X$.

Vector spaces of functions are very important in mathematics and physics, where the field of scalars may be the real numbers, so the ranges of the functions can in turn also be the set of real numbers.

Here's a unique example.

Example VSS The singleton vector space Set: $Z = \{z\}$. Equality: Huh? Vector Addition: z + z = z. Scalar Multiplication: $\alpha z = z$.

This should look pretty wild. First, just what is \mathbf{z} ? Column vector, matrix, polynomial, sequence, function? Mineral, plant, or animal? We aren't saying! \mathbf{z} just *is*. And we have definitions of vector addition and scalar multiplication that are sufficient for an occurrence of either that may come along.

Our only concern is if this set, along with the definitions of two operations, fulfills the ten properties of Definition VS [659]. Let's check associativity of vector addition (Property AA [660]). For all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{Z}$,

$$u + (v + w) = z + (z + z)$$
$$= z + z$$
$$= (z + z) + z$$
$$= (u + v) + w$$

What is the zero vector in this vector space (Property Z [660])? With only one element in the set, we do not have much choice. Is $\mathbf{z} = \mathbf{0}$? It appears that \mathbf{z} behaves like the zero vector should, so it gets the title. Maybe now the definition of this vector space does not seem so

bizarre. It is a set whose only element is the element that behaves like the zero vector, so that lone element is the zero vector.

Perhaps some of the above definitions and verifications seem obvious or like splitting hairs, but the next example should convince you that they *are* necessary. We will study this one carefully. Ready? Check your preconceptions at the door.

Example CVS

The crazy vector space

Set: $C = \{ (x_1, x_2) | x_1, x_2 \in \mathbb{C} \}.$ Vector Addition: $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1 + 1, x_2 + y_2 + 1).$ Scalar Multiplication: $\alpha(x_1, x_2) = (\alpha x_1 + \alpha - 1, \alpha x_2 + \alpha - 1).$

Now, the first thing I hear you say is "You can't do that!" And my response is, "Oh yes, I can!" I am free to define my set and my operations any way I please. They may not look natural, or even useful, but we will now verify that they provide us with another example of a vector space. And that is enough. If you are adventurous, you might try first checking some of the properties yourself. What is the zero vector? Additive inverses? Can you prove associativity? Ready, here we go.

Property AC [660], Property SC [660]: The result of each operation is a pair of complex numbers, so these two closure properties are fulfilled.

Property C [660]:

$$\mathbf{u} + \mathbf{v} = (x_1, x_2) + (y_1, y_2) = (x_1 + y_1 + 1, x_2 + y_2 + 1)$$

= $(y_1 + x_1 + 1, y_2 + x_2 + 1) = (y_1, y_2) + (x_1, x_2)$

$$= \mathbf{v} + \mathbf{u}$$

Property AA [660]:

$$\begin{aligned} \mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (x_1, x_2) + ((y_1, y_2) + (z_1, z_2)) \\ &= (x_1, x_2) + (y_1 + z_1 + 1, y_2 + z_2 + 1) \\ &= (x_1 + (y_1 + z_1 + 1) + 1, x_2 + (y_2 + z_2 + 1) + 1) \\ &= (x_1 + y_1 + z_1 + 2, x_2 + y_2 + z_2 + 2) \\ &= ((x_1 + y_1 + 1) + z_1 + 1, (x_2 + y_2 + 1) + z_2 + 1) \\ &= (x_1 + y_1 + 1, x_2 + y_2 + 1) + (z_1, z_2) \\ &= ((x_1, x_2) + (y_1, y_2)) + (z_1, z_2) \\ &= (\mathbf{u} + \mathbf{v}) + \mathbf{w} \end{aligned}$$

Property Z [660]: The zero vector is $\dots \mathbf{0} = (-1, -1)$. Now I hear you say, "No, no, that can't be, it must be (0, 0)!". Indulge me for a moment and let us check my proposal.

$$\mathbf{u} + \mathbf{0} = (x_1, x_2) + (-1, -1) = (x_1 + (-1) + 1, x_2 + (-1) + 1) = (x_1, x_2) = \mathbf{u}$$

Feeling better? Or worse?

Property AI [660]: For each vector, **u**, we must locate an additive inverse, $-\mathbf{u}$. Here it is, $-(x_1, x_2) = (-x_1 - 2, -x_2 - 2)$. As odd as it may look, I hope you are withholding judgment. Check:

$$\mathbf{u} + (-\mathbf{u}) = (x_1, x_2) + (-x_1 - 2, -x_2 - 2) = (x_1 + (-x_1 - 2) + 1, -x_2 + (x_2 - 2) + 1) = (-1, -1) = \mathbf{0}$$

Property SMA [660]:

$$\begin{aligned} \alpha(\beta \mathbf{u}) &= \alpha(\beta(x_1, x_2)) \\ &= \alpha(\beta x_1 + \beta - 1, \beta x_2 + \beta - 1) \\ &= (\alpha(\beta x_1 + \beta - 1) + \alpha - 1, \alpha(\beta x_2 + \beta - 1) + \alpha - 1) \\ &= ((\alpha \beta x_1 + \alpha \beta - \alpha) + \alpha - 1, (\alpha \beta x_2 + \alpha \beta - \alpha) + \alpha - 1) \\ &= (\alpha \beta x_1 + \alpha \beta - 1, \alpha \beta x_2 + \alpha \beta - 1) \\ &= (\alpha \beta)(x_1, x_2) \\ &= (\alpha \beta) \mathbf{u} \end{aligned}$$

Property DVA [660]: If you have hung on so far, here's where it gets even wilder. In the next two properties we mix and mash the two operations.

$$\begin{aligned} \alpha(\mathbf{u} + \mathbf{v}) &= \alpha \left((x_1, x_2) + (y_1, y_2) \right) \\ &= \alpha(x_1 + y_1 + 1, x_2 + y_2 + 1) \\ &= \left(\alpha(x_1 + y_1 + 1) + \alpha - 1, \alpha(x_2 + y_2 + 1) + \alpha - 1 \right) \\ &= \left(\alpha x_1 + \alpha y_1 + \alpha + \alpha - 1, \alpha x_2 + \alpha y_2 + \alpha + \alpha - 1 \right) \\ &= \left(\alpha x_1 + \alpha - 1 + \alpha y_1 + \alpha - 1 + 1, \alpha x_2 + \alpha - 1 + \alpha y_2 + \alpha - 1 + 1 \right) \\ &= \left((\alpha x_1 + \alpha - 1) + (\alpha y_1 + \alpha - 1) + 1, (\alpha x_2 + \alpha - 1) + (\alpha y_2 + \alpha - 1) + 1 \right) \\ &= \left(\alpha x_1 + \alpha - 1, \alpha x_2 + \alpha - 1 \right) + \left(\alpha y_1 + \alpha - 1, \alpha y_2 + \alpha - 1 \right) \\ &= \alpha(x_1, x_2) + \alpha(y_1, y_2) \end{aligned}$$

 $= \alpha \mathbf{u} + \alpha \mathbf{v}$

Property DSA [660]:

$$\begin{aligned} (\alpha + \beta)\mathbf{u} &= (\alpha + \beta)(x_1, x_2) \\ &= ((\alpha + \beta)x_1 + (\alpha + \beta) - 1, (\alpha + \beta)x_2 + (\alpha + \beta) - 1) \\ &= (\alpha x_1 + \beta x_1 + \alpha + \beta - 1, \alpha x_2 + \beta x_2 + \alpha + \beta - 1) \\ &= (\alpha x_1 + \alpha - 1 + \beta x_1 + \beta - 1 + 1, \alpha x_2 + \alpha - 1 + \beta x_2 + \beta - 1 + 1) \\ &= ((\alpha x_1 + \alpha - 1) + (\beta x_1 + \beta - 1) + 1, (\alpha x_2 + \alpha - 1) + (\beta x_2 + \beta - 1) + 1) \\ &= (\alpha x_1 + \alpha - 1, \alpha x_2 + \alpha - 1) + (\beta x_1 + \beta - 1, \beta x_2 + \beta - 1) \\ &= \alpha (x_1, x_2) + \beta (x_1, x_2) \\ &= \alpha \mathbf{u} + \beta \mathbf{u} \end{aligned}$$

Property O [661]: After all that, this one is easy, but no less pleasing.

$$1\mathbf{u} = 1(x_1, x_2) = (x_1 + 1 - 1, x_2 + 1 - 1) = (x_1, x_2) = \mathbf{u}$$

That's it, C is a vector space, as crazy as that may seem.

Notice that in the case of the zero vector and additive inverses, we only had to propose possibilities and then verify that they were the correct choices. You might try to discover how you would arrive at these choices, though you should understand why the process of discovering them is not a necessary component of the proof itself. \boxtimes

Subsection VSP Vector Space Properties

Subsection VS.EVS [663] has provided us with an abundance of examples of vector spaces, most of them containing useful and interesting mathematical objects along with natural operations. In this subsection we will prove some general properties of vector spaces. Some of these results will again seem obvious, but it is important to understand why it is necessary to state and prove them. A typical hypothesis will be "Let V be a vector space." From this we may assume the ten properties of Definition VS [659], and nothing more. It's like starting over, as we learn about what can happen in this new algebra we are learning. But the power of this careful approach is that we can apply these theorems to any vector space we encounter — those in the previous examples, or new ones we have not yet contemplated. Or perhaps new ones that nobody has ever contemplated. We will illustrate some of these results with examples from the crazy vector space (Example CVS [670]), but mostly we are stating theorems and doing proofs. These proofs do not get too involved, but are not trivial either, so these are good theorems to try proving yourself before you study the proof given here. (See Technique P [1633].)

First we show that there is just one zero vector. Notice that the properties only require there to be *at least* one, and say nothing about there possibly being more. That is because we can use the ten properties of a vector space (Definition VS [659]) to learn that there can *never* be more than one. To require that this extra condition be stated as an eleventh property would make the definition of a vector space more complicated than it needs to be.

Theorem ZVU Zero Vector is Unique

Suppose that V is a vector space. The zero vector, $\mathbf{0}$, is unique.

Proof To prove uniqueness, a standard technique is to suppose the existence of two objects (Technique U [1625]). So let $\mathbf{0}_1$ and $\mathbf{0}_2$ be two zero vectors in V. Then

${f 0}_1={f 0}_1+{f 0}_2$	Property Z [660] for 0_2
$= 0_2 + 0_1$	Property C [660]
$= 0_2$	Property Z [660] for 0_1

This proves the uniqueness since the two zero vectors are really the same.

Theorem AIU Additive Inverses are Unique

Suppose that V is a vector space. For each $\mathbf{u} \in V$, the additive inverse, $-\mathbf{u}$, is unique. \Box

Proof To prove uniqueness, a standard technique is to suppose the existence of two objects (Technique U [1625]). So let $-\mathbf{u}_1$ and $-\mathbf{u}_2$ be two additive inverses for \mathbf{u} . Then

$-\mathbf{u}_1=-\mathbf{u}_1+0$	Property Z $[660]$
$= -\mathbf{u}_1 + (\mathbf{u} + -\mathbf{u}_2)$	Property AI [660]
$=(-\mathbf{u}_1+\mathbf{u})+-\mathbf{u}_2$	Property AA [660]

 \square

$$= \mathbf{0} + -\mathbf{u}_2$$
Property AI [660]
$$= -\mathbf{u}_2$$
Property Z [660]

So the two additive inverses are really the same.

As obvious as the next three theorems appear, nowhere have we guaranteed that the zero scalar, scalar multiplication and the zero vector all interact this way. Until we have proved it, anyway.

Theorem ZSSM Zero Scalar in Scalar Multiplication

Suppose that V is a vector space and $\mathbf{u} \in V$. Then $0\mathbf{u} = \mathbf{0}$.

Proof Notice that 0 is a scalar, **u** is a vector, so Property SC [660] says 0**u** is again a vector. As such, 0**u** has an additive inverse, $-(0\mathbf{u})$ by Property AI [660].

$0\mathbf{u} = 0 + 0\mathbf{u}$	Property Z $[660]$
$= (-(0\mathbf{u}) + 0\mathbf{u}) + 0\mathbf{u}$	Property AI [660]
$= -(0\mathbf{u}) + (0\mathbf{u} + 0\mathbf{u})$	Property AA [660]
$= -(0\mathbf{u}) + (0+0)\mathbf{u}$	Property DSA [660]
$= -(0\mathbf{u}) + 0\mathbf{u}$	Property ZCN [1597]
= 0	Property AI [660]

Here's another theorem that *looks* like it should be obvious, but is still in need of a proof.

Theorem ZVSM Zero Vector in Scalar Multiplication

Suppose that V is a vector space and $\alpha \in \mathbb{C}$. Then $\alpha \mathbf{0} = \mathbf{0}$.

Proof Notice that α is a scalar, **0** is a vector, so Property SC [660] means α **0** is again a vector. As such, α **0** has an additive inverse, $-(\alpha$ **0**) by Property AI [660].

$\alpha 0 = 0 + \alpha 0$	Property Z $[660]$
$= (-(\alpha 0) + \alpha 0) + \alpha 0$	Property AI [660]
$= -(\alpha 0) + (\alpha 0 + \alpha 0)$	Property AA [660]
$= -(\alpha 0) + \alpha \left(0 + 0\right)$	Property DVA [660]
$= -(\alpha 0) + \alpha 0$	Property Z $[660]$
= 0	Property AI [660]

Here's another one that sure looks obvious. But understand that we have chosen to use certain notation because it makes the theorem's conclusion look so nice. The theorem is not true because the notation looks so good, it still needs a proof. If we had really wanted to make this point, we might have defined the additive inverse of \mathbf{u} as \mathbf{u}^{\sharp} . Then we would have written the defining property, Property AI [660], as $\mathbf{u} + \mathbf{u}^{\sharp} = \mathbf{0}$. This theorem would become $\mathbf{u}^{\sharp} = (-1)\mathbf{u}$. Not really quite as pretty, is it?

Theorem AISM

Additive Inverses from Scalar Multiplication

Suppose that V is a vector space and $\mathbf{u} \in V$. Then $-\mathbf{u} = (-1)\mathbf{u}$.

\mathbf{Proof}

$-\mathbf{u} = -\mathbf{u} + 0$	Property Z $[660]$
$= -\mathbf{u} + 0\mathbf{u}$	Theorem ZSSM [676]
$= -\mathbf{u} + (1 + (-1))\mathbf{u}$	
$= -\mathbf{u} + (1\mathbf{u} + (-1)\mathbf{u})$	Property DSA [660]
$= -\mathbf{u} + (\mathbf{u} + (-1)\mathbf{u})$	Property O [661]
$= (-\mathbf{u} + \mathbf{u}) + (-1)\mathbf{u}$	Property AA [660]
$= 0 + (-1)\mathbf{u}$	Property AI [660]
$=(-1)\mathbf{u}$	Property Z $[660]$

Because of this theorem, we can now write linear combinations like $6\mathbf{u}_1 + (-4)\mathbf{u}_2$ as $6\mathbf{u}_1 - 4\mathbf{u}_2$, even though we have not formally defined an operation called **vector subtraction**. Our next theorem is a bit different from several of the others in the list. Rather than making a declaration ("the zero vector is unique") it is an implication ("if..., then...") and so can be used in proofs to convert a vector equality into two possibilities, one a scalar equality and the other a vector equality. It should remind you of the situation for complex numbers. If α , $\beta \in \mathbb{C}$ and $\alpha\beta = 0$, then $\alpha = 0$ or $\beta = 0$. This critical property is the driving force behind using a factorization to solve a polynomial equation.

 \square

Theorem SMEZV Scalar Multiplication Equals the Zero Vector

Suppose that V is a vector space and $\alpha \in \mathbb{C}$. If $\alpha \mathbf{u} = \mathbf{0}$, then either $\alpha = 0$ or $\mathbf{u} = \mathbf{0}$.

Proof We prove this theorem by breaking up the analysis into two cases. The first seems too trivial, and it is, but the logic of the argument is still legitimate.

Case 1. Suppose $\alpha = 0$. In this case our conclusion is true (the first part of the either/or is true) and we are done. That was easy.

Case 2. Suppose $\alpha \neq 0$.

$\mathbf{u} = 1\mathbf{u}$	Property O [661]
$=\left(\frac{1}{lpha}lpha ight)\mathbf{u}$	$\alpha \neq 0$
$=rac{1}{lpha}\left(lpha\mathbf{u} ight)$	Property SMA [660]
$=rac{1}{lpha}\left(0 ight)$	Hypothesis
= 0	Theorem ZVSM [677]

So in this case, the conclusion is true (the second part of the either/or is true) and we are done since the conclusion was true in each of the two cases.

Example PCVS Properties for the Crazy Vector Space

Several of the above theorems have interesting demonstrations when applied to the crazy

 \square

vector space, C (Example CVS [670]). We are not proving anything new here, or learning anything we did not know already about C. It is just plain fun to see how these general theorems apply in a specific instance. For most of our examples, the applications are obvious or trivial, but not with C.

Suppose $\mathbf{u} \in C$. Then, as given by Theorem ZSSM [676],

$$0\mathbf{u} = 0(x_1, x_2) = (0x_1 + 0 - 1, 0x_2 + 0 - 1) = (-1, -1) = \mathbf{0}$$

And as given by Theorem ZVSM [677],

$$\alpha \mathbf{0} = \alpha(-1, -1) = (\alpha(-1) + \alpha - 1, \alpha(-1) + \alpha - 1)$$

= $(-\alpha + \alpha - 1, -\alpha + \alpha - 1) = (-1, -1) = \mathbf{0}$

Finally, as given by Theorem AISM [678],

$$(-1)\mathbf{u} = (-1)(x_1, x_2) = ((-1)x_1 + (-1) - 1, (-1)x_2 + (-1) - 1)$$
$$= (-x_1 - 2, -x_2 - 2) = -\mathbf{u}$$

 \boxtimes

Subsection RD Recycling Definitions

When we say that V is a vector space, we then know we have a set of objects (the "vectors"), but we also know we have been provided with two operations ("vector addition" and "scalar multiplication") and these operations behave with these objects according to the ten properties of Definition VS [659]. One combines two vectors and produces a vector, the other takes a scalar and a vector, producing a vector as the result. So if $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \in V$ then an expression like

$5\mathbf{u}_1 + 7\mathbf{u}_2 - 13\mathbf{u}_3$

would be unambiguous in *any* of the vector spaces we have discussed in this section. And the resulting object would be another vector in the vector space. If you were tempted to call the above expression a linear combination, you would be right. Four of the definitions that were central to our discussions in Chapter V [197] were stated in the context of vectors being *column vectors*, but were purposely kept broad enough that they could be applied in the context of any vector space. They only rely on the presence of scalars, vectors, vector addition and scalar multiplication to make sense. We will restate them shortly, unchanged, except that their titles and acronyms no longer refer to column vectors, and the hypothesis of being in a vector space has been added. Take the time now to look forward and review each one, and begin to form some connections to what we have done earlier and what we will be doing in subsequent sections and chapters. Specifically, compare the following pairs of definitions:

Definition LCCV [221] and Definition LC [707] Definition SSCV [270] and Definition SS [709] Definition RLDCV [316] and Definition RLD [735] Definition LICV [317] and Definition LI [735]

Subsection READ Reading Questions

- 1. Comment on how the vector space \mathbb{C}^m went from a theorem (Theorem VSPCV [205]) to an example (Example VSCV [663]).
- 2. In the crazy vector space, C, (Example CVS [670]) compute the linear combination

$$2(3, 4) + (-6)(1, 2).$$

3. Suppose that α is a scalar and **0** is the zero vector. Why should we prove anything as obvious as $\alpha \mathbf{0} = \mathbf{0}$ such as we did in Theorem ZVSM [677]?

Subsection EXC Exercises

M10 Define a possibly new vector space by beginning with the set and vector addition from \mathbb{C}^2 (Example VSCV [663]) but change the definition of scalar multiplication to

$$\alpha \mathbf{x} = \mathbf{0} = \begin{bmatrix} 0\\0 \end{bmatrix} \qquad \qquad \alpha \in \mathbb{C}, \ \mathbf{x} \in \mathbb{C}^2$$

Prove that the first nine properties required for a vector space hold, but Property O [661] does not hold.

This example shows us that we cannot expect to be able to derive Property O [661] as a consequence of assuming the first nine properties. In other words, we cannot slim down our list of properties by jettisoning the last one, and still have the same collection of objects qualify as vector spaces.

Contributed by Robert Beezer

M11 Let V be the set \mathbb{C}^2 with the usual vector addition, but with scalar multiplication defined by

$$\alpha \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \alpha y \\ \alpha x \end{bmatrix}$$

Determine whether or not V is a vector space with these operations. Contributed by Chris Black Solution [688] **M12** Let V be the set \mathbb{C}^2 with the usual scalar multiplication, but with vector addition defined by

$$\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} y+w \\ x+z \end{bmatrix}$$

Determine whether or not V is a vector space with these operations. Contributed by Chris Black Solution [688]

M13 Let V be the set $M_{2,2}$ with the usual scalar multiplication, but with addition defined by $A + B = \mathcal{O}_{2,2}$ for all 2×2 matrices A and B. Determine whether or not V is a vector space with these operations.

Contributed by Chris Black Solution [689]

M14 Let V be the set $M_{2,2}$ with the usual addition, but with scalar multiplication defined by $\alpha A = \mathcal{O}_{2,2}$ for all 2×2 matrices A and scalars α . Determine whether or not V is a vector space with these operations.

Contributed by Chris Black Solution [689]

M15 Consider the following sets of 3×3 matrices, where the symbol * indicates the position of an arbitrary complex number. Determine whether or not these sets form vector spaces with the usual operations of addition and scalar multiplication for matrices.

1. All matrices of the form
$$\begin{bmatrix} * & * & 1 \\ * & 1 & * \\ 1 & * & * \end{bmatrix}$$

2. All matrices of the form
$$\begin{bmatrix} * & 0 & * \\ 0 & * & 0 \\ * & 0 & * \end{bmatrix}$$
3. All matrices of the form $\begin{bmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \end{bmatrix}$ (These are the diagonal matrices.)4. All matrices of the form $\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}$ (These are the upper triangular matrices.)

Contributed by Chris Black Solution [689]

M20 Explain why we need to define the vector space P_n as the set of all polynomials with degree up to and including n instead of the more obvious set of all polynomials of degree exactly n.

Contributed by Chris Black Solution [690]

M21 Does the set $\mathbb{Z}^2 = \left\{ \begin{bmatrix} m \\ n \end{bmatrix} \middle| m, n \in \mathbb{Z} \right\}$ with the operations of standard addition and scalar multiplication of vectors form a vector space? Contributed by Chris Black Solution [690]

T10 Prove each of the ten properties of Definition VS [659] for each of the following

examples of a vector space: Example VSP [665] Example VSIS [667] Example VSF [667] Example VSS [668] Contributed by Robert Beezer

The next three problems suggest that under the right situations we can "cancel." In practice, these techniques should be avoided in other proofs. Prove each of the following statements.

T21 Suppose that V is a vector space, and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$. If $\mathbf{w} + \mathbf{u} = \mathbf{w} + \mathbf{v}$, then $\mathbf{u} = \mathbf{v}$. Contributed by Robert Beezer Solution [690]

T22 Suppose V is a vector space, $\mathbf{u}, \mathbf{v} \in V$ and α is a nonzero scalar from \mathbb{C} . If $\alpha \mathbf{u} = \alpha \mathbf{v}$, then $\mathbf{u} = \mathbf{v}$.

Contributed by Robert Beezer Solution [691]

T23 Suppose V is a vector space, $\mathbf{u} \neq \mathbf{0}$ is a vector in V and $\alpha, \beta \in \mathbb{C}$. If $\alpha \mathbf{u} = \beta \mathbf{u}$, then $\alpha = \beta$.

Contributed by Robert Beezer Solution [691]

T30 Suppose that V is a vector space and $\alpha \in \mathbb{C}$ is a scalar such that $\alpha \mathbf{x} = \mathbf{x}$ for every $\mathbf{x} \in V$. Prove that $\alpha = 1$. In other words, Property O [661] is not duplicated for any other scalar but the "special" scalar, 1. (This question was suggested by James Gallagher.)

Contributed by Robert Beezer Solution [692]

Subsection SOL Solutions

M11 Contributed by Chris Black Statement [683] The set \mathbb{C}^2 with the proposed operations is not a vector space since Property O [661] is not valid. A counterexample is $\begin{bmatrix} 3\\2 \end{bmatrix} = \begin{bmatrix} 2\\3 \end{bmatrix} \neq \begin{bmatrix} 3\\2 \end{bmatrix}$, so in general, $\mathbf{1u} \neq \mathbf{u}$. M12 Contributed by Chris Black Statement [684] Let's consider the existence of a zero vector, as required by Property Z [660] of a vector space. The "regular" zero vector fails : $\begin{bmatrix} x\\y \end{bmatrix} + \begin{bmatrix} 0\\0 \end{bmatrix} = \begin{bmatrix} y\\x \end{bmatrix} \neq \begin{bmatrix} x\\y \end{bmatrix}$ (remember that the property must hold for *every* vector, not just for some). Is there another vector that fills the role of the zero vector? Suppose that $\mathbf{0} = \begin{bmatrix} z_1\\z_2 \end{bmatrix}$. Then for any vector $\begin{bmatrix} x\\y \end{bmatrix}$, we have $\begin{bmatrix} x\\y \end{bmatrix} + \begin{bmatrix} z_1\\z_2 \end{bmatrix} = \begin{bmatrix} y+z_2\\x+z_1 \end{bmatrix} = \begin{bmatrix} x\\y \end{bmatrix}$

so that $x = y + z_2$ and $y = x + z_1$. This means that $z_1 = y - x$ and $z_2 = x - y$. However, since x and y can be any complex numbers, there are no fixed complex numbers z_1 and z_2 that satisfy these equations. Thus, there is no zero vector, Property Z [660] is not valid, and the set \mathbb{C}^2 with the proposed operations is not a vector space.

M13 Contributed by Chris Black Statement [684]

Since scalar multiplication remains unchanged, we only need to consider the axioms that involve vector addition. Since every sum is the zero matrix, the first 4 properties hold easily. However, there is no zero vector in this set. Suppose that there was. Then there is a matrix Z so that A + Z = A for any 2×2 matrix A. However, $A + Z = \mathcal{O}_{2,2}$, which is in general not equal to A, so Property Z [660] fails and this set is not a vector space.

M14 Contributed by Chris Black Statement [684]

Since addition is unchanged, we only need to check the axioms involving scalar multiplication. The proposed scalar multiplication clearly fails Property O [661] : $1A = \mathcal{O}_{2,2} \neq A$. Thus, the proposed set is not a vector space.

M15 Contributed by Chris Black Statement [684]

There is something to notice here that will make our job much easier: Since each of these sets are comprised of 3×3 matrices with the standard operations of addition and scalar multiplication of matrices, the last 8 properties will automatically hold. That is, we really only need to verify Property AC [660] and Property SC [660].

a). This set is not closed under either scalar multiplication or addition (fails Property AC

[660] and Property SC [660]). For example, $3\begin{bmatrix} * & * & 1 \\ * & 1 & * \\ 1 & * & * \end{bmatrix} = \begin{bmatrix} * & * & 3 \\ * & 3 & * \\ 3 & * & * \end{bmatrix}$ is not a member of the proposed set

of the proposed set.

b). This set is closed under both scalar multiplication and addition, so this set is a vector

space with the standard operation of addition and scalar multiplication.

- c). This set is closed under both scalar multiplication and addition, so this set is a vector space with the standard operation of addition and scalar multiplication.
- d). This set is closed under both scalar multiplication and addition, so this set is a vector space with the standard operation of addition and scalar multiplication.

M20 Contributed by Chris Black Statement [685] Hint: The set of all polynomials of degree *exactly* n fails one of the closure properties of a vector space. Which one, and why?

M21 Contributed by Robert Beezer Statement [685] Additive closure will hold, but scalar closure will not. The best way to convince yourself of this is to construct a counterexample. Such as, $\frac{1}{2} \in \mathbb{C}$ and $\begin{bmatrix} 1\\0 \end{bmatrix} \in \mathbb{Z}^2$, however $\frac{1}{2} \begin{bmatrix} 1\\0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\\0 \end{bmatrix} \notin \mathbb{Z}^2$, which violates Property SC [660]. So \mathbb{Z}^2 is not a vector space.

T21 Contributed by Robert Beezer Statement [686]

$\mathbf{u} = 0 + \mathbf{u}$	Property Z $[660]$
$= (-\mathbf{w} + \mathbf{w}) + \mathbf{u}$	Property AI [660]
$= -\mathbf{w} + (\mathbf{w} + \mathbf{u})$	Property AA [660]
$= -\mathbf{w} + (\mathbf{w} + \mathbf{v})$	Hypothesis

$$= (-\mathbf{w} + \mathbf{w}) + \mathbf{v}$$
Property AA [660]
$$= \mathbf{0} + \mathbf{v}$$
Property AI [660]
$$= \mathbf{v}$$
Property Z [660]

T22 Contributed by Robert Beezer Statement [686]

 $\mathbf{u} = 1\mathbf{u}$ $= \left(\frac{1}{\alpha}\alpha\right)\mathbf{u}$ $= \frac{1}{\alpha}\left(\alpha\mathbf{u}\right)$ $= \frac{1}{\alpha}\left(\alpha\mathbf{v}\right)$ $= \left(\frac{1}{\alpha}\alpha\right)\mathbf{v}$ $= 1\mathbf{v}$ $= \mathbf{v}$

Property O [661]

 $\alpha \neq 0$

Property SMA [660]

Hypothesis

Property SMA [660]

Property O [661]

T23 Contributed by Robert Beezer Statement [686]

 $\mathbf{0} = \alpha \mathbf{u} + -(\alpha \mathbf{u}) \qquad \qquad \text{Property AI [660]}$

$$= \beta \mathbf{u} + -(\alpha \mathbf{u})$$
Hypothesis $= \beta \mathbf{u} + (-1) (\alpha \mathbf{u})$ Theorem AISM [678] $= \beta \mathbf{u} + ((-1)\alpha) \mathbf{u}$ Property SMA [660] $= \beta \mathbf{u} + (-\alpha) \mathbf{u}$ Property DSA [660]

By hypothesis, $\mathbf{u} \neq \mathbf{0}$, so Theorem SMEZV [679] implies

$$0 = \beta - \alpha$$
$$\alpha = \beta$$

T30 Contributed by Robert Beezer Statement [686] We have,

$0 = \mathbf{x} - \mathbf{x}$	Property AI [660]
$= \alpha \mathbf{x} - \mathbf{x}$	Hypothesis
$= \alpha \mathbf{x} - 1 \mathbf{x}$	Property O [661]
$= (\alpha - 1)\mathbf{x}$	Property DSA [660]

So by Theorem SMEZV [679] we conclude that $\alpha - 1 = 0$ or $\mathbf{x} = \mathbf{0}$. However, since our hypothesis was for every $\mathbf{x} \in V$, we are left with the first possibility and $\alpha = 1$.

There is one flaw in the proof above, and as stated, the problem is not correct either. Can you spot the flaw and as a result correct the problem statement? (Hint: Example VSS [668]).

Section S Subspaces

A subspace is a vector space that is contained within another vector space. So every subspace is a vector space in its own right, but it is also defined relative to some other (larger) vector space. We will discover shortly that we are already familiar with a wide variety of subspaces from previous sections. Here's the definition.

Definition S

Subspace

Suppose that V and W are two vector spaces that have identical definitions of vector addition and scalar multiplication, and that W is a subset of $V, W \subseteq V$. Then W is a **subspace** of V. \triangle

Lets look at an example of a vector space inside another vector space.

Example SC3

A subspace of \mathbb{C}^3

We know that \mathbb{C}^3 is a vector space (Example VSCV [663]). Consider the subset,

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \middle| 2x_1 - 5x_2 + 7x_3 = 0 \right\}$$

It is clear that $W \subseteq \mathbb{C}^3$, since the objects in W are column vectors of size 3. But is W a vector space? Does it satisfy the ten properties of Definition VS [659] when we use the same operations? That is the main question. Suppose $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ are vectors from

W. Then we know that these vectors cannot be totally arbitrary, they must have gained membership in W by virtue of meeting the membership test. For example, we know that **x** must satisfy $2x_1 - 5x_2 + 7x_3 = 0$ while **y** must satisfy $2y_1 - 5y_2 + 7y_3 = 0$. Our first property (Property AC [660]) asks the question, is $\mathbf{x} + \mathbf{y} \in W$? When our set of vectors was \mathbb{C}^3 , this was an easy question to answer. Now it is not so obvious. Notice first that

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}$$

and we can test this vector for membership in W as follows,

$$2(x_1 + y_1) - 5(x_2 + y_2) + 7(x_3 + y_3) = 2x_1 + 2y_1 - 5x_2 - 5y_2 + 7x_3 + 7y_3$$

= $(2x_1 - 5x_2 + 7x_3) + (2y_1 - 5y_2 + 7y_3)$
= $0 + 0$
= 0
x $\in W, y \in W$
= 0

and by this computation we see that $\mathbf{x} + \mathbf{y} \in W$. One property down, nine to go.

If α is a scalar and $\mathbf{x} \in W$, is it always true that $\alpha \mathbf{x} \in W$? This is what we need to establish Property SC [660]. Again, the answer is not as obvious as it was when our set of vectors was all of \mathbb{C}^3 . Let's see.

$$\alpha \mathbf{x} = \alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{bmatrix}$$

and we can test this vector for membership in W with

$$2(\alpha x_1) - 5(\alpha x_2) + 7(\alpha x_3) = \alpha(2x_1 - 5x_2 + 7x_3)$$
$$= \alpha 0$$
$$\mathbf{x} \in W$$
$$= 0$$

and we see that indeed $\alpha \mathbf{x} \in W$. Always.

If W has a zero vector, it will be unique (Theorem ZVU [675]). The zero vector for \mathbb{C}^3 should also perform the required duties when added to elements of W. So the likely candidate for a zero vector in W is the same zero vector that we know \mathbb{C}^3 has. You can check that $\mathbf{0} = \begin{bmatrix} 0\\0\\0 \end{bmatrix}$ is a zero vector in W too (Property Z [660]).

With a zero vector, we can now ask about additive inverses (Property AI [660]). As you might suspect, the natural candidate for an additive inverse in W is the same as the additive inverse from \mathbb{C}^3 . However, we must insure that these additive inverses actually are elements

of W. Given $\mathbf{x} \in W$, is $-\mathbf{x} \in W$?

$$-\mathbf{x} = \begin{bmatrix} -x_1 \\ -x_2 \\ -x_3 \end{bmatrix}$$

and we can test this vector for membership in W with

$$2(-x_1) - 5(-x_2) + 7(-x_3) = -(2x_1 - 5x_2 + 7x_3)$$

= -0 $\mathbf{x} \in W$
= 0

and we now believe that $-\mathbf{x} \in W$.

Is the vector addition in W commutative (Property C [660])? Is $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$? Of course! Nothing about restricting the scope of our set of vectors will prevent the operation from still being commutative. Indeed, the remaining five properties are unaffected by the transition to a smaller set of vectors, and so remain true. That was convenient.

So W satisfies all ten properties, is therefore a vector space, and thus earns the title of being a subspace of \mathbb{C}^3 .

Subsection TS Testing Subspaces

In Example SC3 [694] we proceeded through all ten of the vector space properties before believing that a subset was a subspace. But six of the properties were easy to prove, and we can lean on some of the properties of the vector space (the superset) to make the other four easier. Here is a theorem that will make it easier to test if a subset is a vector space. A shortcut if there ever was one.

Theorem TSS

Testing Subsets for Subspaces

Suppose that V is a vector space and W is a subset of $V, W \subseteq V$. Endow W with the same operations as V. Then W is a subspace if and only if three conditions are met

- 1. W is non-empty, $W \neq \emptyset$.
- 2. If $\mathbf{x} \in W$ and $\mathbf{y} \in W$, then $\mathbf{x} + \mathbf{y} \in W$.
- 3. If $\alpha \in \mathbb{C}$ and $\mathbf{x} \in W$, then $\alpha \mathbf{x} \in W$.

Proof (\Rightarrow) We have the hypothesis that W is a subspace, so by Property Z [660] we know that W contains a zero vector. This is enough to show that $W \neq \emptyset$. Also, since W is a

vector space it satisfies the additive and scalar multiplication closure properties (Property AC [660], Property SC [660]), and so exactly meets the second and third conditions. If that was easy, the other direction might require a bit more work.

(\Leftarrow) We have three properties for our hypothesis, and from this we should conclude that W has the ten defining properties of a vector space. The second and third conditions of our hypothesis are exactly Property AC [660] and Property SC [660]. Our hypothesis that V is a vector space implies that Property C [660], Property AA [660], Property SMA [660], Property DVA [660], Property DSA [660] and Property O [661] all hold. They continue to be true for vectors from W since passing to a subset, and keeping the operation the same, leaves their statements unchanged. Eight down, two to go.

Suppose $\mathbf{x} \in W$. Then by the third part of our hypothesis (scalar closure), we know that $(-1)\mathbf{x} \in W$. By Theorem AISM [678] $(-1)\mathbf{x} = -\mathbf{x}$, so together these statements show us that $-\mathbf{x} \in W$. $-\mathbf{x}$ is the additive inverse of \mathbf{x} in V, but will continue in this role when viewed as element of the subset W. So every element of W has an additive inverse that is an element of W and Property AI [660] is completed. Just one property left.

While we have implicitly discussed the zero vector in the previous paragraph, we need to be certain that the zero vector (of V) really lives in W. Since W is non-empty, we can choose some vector $\mathbf{z} \in W$. Then by the argument in the previous paragraph, we know $-\mathbf{z} \in W$. Now by Property AI [660] for V and then by the second part of our hypothesis (additive closure) we see that

$$\mathbf{0} = \mathbf{z} + (-\mathbf{z}) \in W$$

So W contain the zero vector from V. Since this vector performs the required duties of a

zero vector in V, it will continue in that role as an element of W. This gives us, Property Z [660], the final property of the ten required. (Sarah Fellez contributed to this proof.)

So just three conditions, plus being a subset of a known vector space, gets us all ten properties. Fabulous! This theorem can be paraphrased by saying that a subspace is "a non-empty subset (of a vector space) that is closed under vector addition and scalar multiplication."

You might want to go back and rework Example SC3 [694] in light of this result, perhaps seeing where we can now economize or where the work done in the example mirrored the proof and where it did not. We will press on and apply this theorem in a slightly more abstract setting.

Example SP4

A subspace of P_4

 P_4 is the vector space of polynomials with degree at most 4 (Example VSP [665]). Define a subset W as

$$W = \{ p(x) | p \in P_4, \ p(2) = 0 \}$$

so W is the collection of those polynomials (with degree 4 or less) whose graphs cross the x-axis at x = 2. Whenever we encounter a new set it is a good idea to gain a better understanding of the set by finding a few elements in the set, and a few outside it. For example $x^2 - x - 2 \in W$, while $x^4 + x^3 - 7 \notin W$.

Is W nonempty? Yes, $x - 2 \in W$.

Additive closure? Suppose $p \in W$ and $q \in W$. Is $p + q \in W$? p and q are not totally arbitrary, we know that p(2) = 0 and q(2) = 0. Then we can check p + q for membership in W,

$$(p+q)(2) = p(2) + q(2)$$

Addition in P_4
$$= 0 + 0$$

$$p \in W, q \in W$$

$$= 0$$

so we see that p + q qualifies for membership in W.

Scalar multiplication closure? Suppose that $\alpha \in \mathbb{C}$ and $p \in W$. Then we know that p(2) = 0. Testing αp for membership,

 $(\alpha p)(2) = \alpha p(2) \qquad Scalar multiplication in P_4$ $= \alpha 0 \qquad p \in W$ = 0

so $\alpha p \in W$.

We have shown that W meets the three conditions of Theorem TSS [698] and so qualifies as a subspace of P_4 . Notice that by Definition S [694] we now know that W is also a vector space. So all the properties of a vector space (Definition VS [659]) and the theorems of Section VS [659] apply in full.

\boxtimes

Much of the power of Theorem TSS [698] is that we can easily establish new vector spaces if we can locate them as subsets of other vector spaces, such as the ones presented in Subsection VS.EVS [663].

It can be as instructive to consider some subsets that are *not* subspaces. Since Theorem TSS [698] is an equivalence (see Technique E [1619]) we can be assured that a subset is not a subspace if it violates one of the three conditions, and in any example of interest this will not be the "non-empty" condition. However, since a subspace has to be a vector space in its own right, we can also search for a violation of any one of the ten defining properties in Definition VS [659] or any inherent property of a vector space, such as those given by the basic theorems of Subsection VS.VSP [674]. Notice also that a violation need only be for a specific vector or pair of vectors.

Example NSC2Z A non-subspace in \mathbb{C}^2 , zero vector Consider the subset W below as a candidate for being a subspace of \mathbb{C}^2

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| 3x_1 - 5x_2 = 12 \right\}$$

The zero vector of \mathbb{C}^2 , $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ will need to be the zero vector in W also. However, $\mathbf{0} \notin W$ since $3(0) - 5(0) = 0 \neq 12$. So W has no zero vector and fails Property Z [660] of Definition VS [659]. This subspace also fails to be closed under addition and scalar multiplication. Can

you find examples of this?

Example NSC2A A non-subspace in \mathbb{C}^2 , additive closure

Consider the subset X below as a candidate for being a subspace of \mathbb{C}^2

$$X = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1 x_2 = 0 \right\}$$

You can check that $\mathbf{0} \in X$, so the approach of the last example will not get us anywhere. However, notice that $\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in X$ and $\mathbf{y} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in X$. Yet

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} 1\\ 0 \end{bmatrix} + \begin{bmatrix} 0\\ 1 \end{bmatrix} = \begin{bmatrix} 1\\ 1 \end{bmatrix} \notin X$$

So X fails the additive closure requirement of either Property AC [660] or Theorem TSS [698], and is therefore not a subspace. \boxtimes

Example NSC2S

A non-subspace in \mathbb{C}^2 , scalar multiplication closure Consider the subset Y below as a candidate for being a subspace of \mathbb{C}^2

$$Y = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1 \in \mathbb{Z}, \, x_2 \in \mathbb{Z} \right\}$$

 \boxtimes

 \mathbb{Z} is the set of integers, so we are only allowing "whole numbers" as the constituents of our vectors. Now, $\mathbf{0} \in Y$, and additive closure also holds (can you prove these claims?). So we will have to try something different. Note that $\alpha = \frac{1}{2} \in \mathbb{C}$ and $\begin{bmatrix} 2\\ 3 \end{bmatrix} \in Y$, but

$$\alpha \mathbf{x} = \frac{1}{2} \begin{bmatrix} 2\\ 3 \end{bmatrix} = \begin{bmatrix} 1\\ \frac{3}{2} \end{bmatrix} \notin Y$$

So Y fails the scalar multiplication closure requirement of either Property SC [660] or Theorem TSS [698], and is therefore not a subspace. \boxtimes

There are two examples of subspaces that are trivial. Suppose that V is any vector space. Then V is a subset of itself and is a vector space. By Definition S [694], V qualifies as a subspace of itself. The set containing just the zero vector $Z = \{0\}$ is also a subspace as can be seen by applying Theorem TSS [698] or by simple modifications of the techniques hinted at in Example VSS [668]. Since these subspaces are so obvious (and therefore not too interesting) we will refer to them as being trivial.

Definition TS Trivial Subspaces

Given the vector space V, the subspaces V and $\{0\}$ are each called a **trivial subspace**. \triangle

We can also use Theorem TSS [698] to prove more general statements about subspaces, as illustrated in the next theorem.

Theorem NSMS

Null Space of a Matrix is a Subspace

Suppose that A is an $m \times n$ matrix. Then the null space of $A, \mathcal{N}(A)$, is a subspace of \mathbb{C}^n . \Box

Proof We will examine the three requirements of Theorem TSS [698]. Recall that $\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{C}^n | A\mathbf{x} = \mathbf{0}\}.$

First, $\mathbf{0} \in \mathcal{N}(A)$, which can be inferred as a consequence of Theorem HSC [147]. So $\mathcal{N}(A) \neq \emptyset$.

Second, check additive closure by supposing that $\mathbf{x} \in \mathcal{N}(A)$ and $\mathbf{y} \in \mathcal{N}(A)$. So we know a little something about \mathbf{x} and \mathbf{y} : $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{y} = \mathbf{0}$, and that is all we know. Question: Is $\mathbf{x} + \mathbf{y} \in \mathcal{N}(A)$? Let's check.

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$$
Theorem MMDAA [476] $= \mathbf{0} + \mathbf{0}$ $\mathbf{x} \in \mathcal{N}(A), \ \mathbf{y} \in \mathcal{N}(A)$ $= \mathbf{0}$ Theorem VSPCV [205]

So, yes, $\mathbf{x} + \mathbf{y}$ qualifies for membership in $\mathcal{N}(A)$.

Third, check scalar multiplication closure by supposing that $\alpha \in \mathbb{C}$ and $\mathbf{x} \in \mathcal{N}(A)$. So we know a little something about \mathbf{x} : $A\mathbf{x} = \mathbf{0}$, and that is all we know. Question: Is $\alpha \mathbf{x} \in \mathcal{N}(A)$? Let's check.

$A(\alpha \mathbf{x}) = \alpha(A\mathbf{x})$	Theorem MMSMM [477]
$= \alpha 0$	$\mathbf{x} \in \mathcal{N}(A)$
= 0	Theorem ZVSM [677]

So, yes, $\alpha \mathbf{x}$ qualifies for membership in $\mathcal{N}(A)$.

Having met the three conditions in Theorem TSS [698] we can now say that the null space of a matrix is a subspace (and hence a vector space in its own right!). \blacksquare

Here is an example where we can exercise Theorem NSMS [705].

Example RSNS Recasting a subspace as a null space

Consider the subset of \mathbb{C}^5 defined as

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \middle| \begin{array}{c} 3x_1 + x_2 - 5x_3 + 7x_4 + x_5 = 0, \\ 4x_1 + 6x_2 + 3x_3 - 6x_4 - 5x_5 = 0, \\ -2x_1 + 4x_2 + 7x_4 + x_5 = 0 \end{array} \right\}$$

It is possible to show that W is a subspace of \mathbb{C}^5 by checking the three conditions of Theorem TSS [698] directly, but it will get tedious rather quickly. Instead, give W a fresh look and notice that it is a set of solutions to a homogeneous system of equations. Define the matrix

$$A = \begin{bmatrix} 3 & 1 & -5 & 7 & 1 \\ 4 & 6 & 3 & -6 & -5 \\ -2 & 4 & 0 & 7 & 1 \end{bmatrix}$$

and then recognize that $W = \mathcal{N}(A)$. By Theorem NSMS [705] we can immediately see that W is a subspace. Boom!

Subsection TSS The Span of a Set

The span of a set of column vectors got a heavy workout in Chapter V [197] and Chapter M [425]. The definition of the span depended only on being able to formulate linear combinations. In any of our more general vector spaces we always have a definition of vector addition and of scalar multiplication. So we can build linear combinations and manufacture spans. This subsection contains two definitions that are just mild variants of definitions we have seen earlier for column vectors. If you haven't already, compare them with Definition LCCV [221] and Definition SSCV [270].

Definition LC Linear Combination

Suppose that V is a vector space. Given n vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n$ and n scalars $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha$ their **linear combination** is the vector

$$\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2+\alpha_3\mathbf{u}_3+\cdots+\alpha_n\mathbf{u}_n.$$

 \triangle

Example LCM A linear combination of matrices In the vector space M_{23} of 2×3 matrices, we have the vectors

$$\mathbf{x} = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 0 & 7 \end{bmatrix} \qquad \qquad \mathbf{y} = \begin{bmatrix} 3 & -1 & 2 \\ 5 & 5 & 1 \end{bmatrix} \qquad \qquad \mathbf{z} = \begin{bmatrix} 4 & 2 & -4 \\ 1 & 1 & 1 \end{bmatrix}$$

and we can form linear combinations such as

$$2\mathbf{x} + 4\mathbf{y} + (-1)\mathbf{z} = 2\begin{bmatrix} 1 & 3 & -2\\ 2 & 0 & 7 \end{bmatrix} + 4\begin{bmatrix} 3 & -1 & 2\\ 5 & 5 & 1 \end{bmatrix} + (-1)\begin{bmatrix} 4 & 2 & -4\\ 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 2 & 6 & -4\\ 4 & 0 & 14 \end{bmatrix} + \begin{bmatrix} 12 & -4 & 8\\ 20 & 20 & 4 \end{bmatrix} + \begin{bmatrix} -4 & -2 & 4\\ -1 & -1 & -1 \end{bmatrix}$$
$$= \begin{bmatrix} 10 & 0 & 8\\ 23 & 19 & 17 \end{bmatrix}$$

or,

$$4\mathbf{x} - 2\mathbf{y} + 3\mathbf{z} = 4\begin{bmatrix} 1 & 3 & -2\\ 2 & 0 & 7 \end{bmatrix} - 2\begin{bmatrix} 3 & -1 & 2\\ 5 & 5 & 1 \end{bmatrix} + 3\begin{bmatrix} 4 & 2 & -4\\ 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 4 & 12 & -8\\ 8 & 0 & 28 \end{bmatrix} + \begin{bmatrix} -6 & 2 & -4\\ -10 & -10 & -2 \end{bmatrix} + \begin{bmatrix} 12 & 6 & -12\\ 3 & 3 & 3 \end{bmatrix}$$
$$= \begin{bmatrix} 10 & 20 & -24\\ 1 & -7 & 29 \end{bmatrix}$$

 \boxtimes

When we realize that we can form linear combinations in any vector space, then it is natural to revisit our definition of the span of a set, since it is the set of *all* possible linear combinations of a set of vectors.

Definition SS Span of a Set

Suppose that V is a vector space. Given a set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t}$, their **span**, $\langle S \rangle$, is the set of all possible linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t$. Symbolically,

$$\langle S \rangle = \left\{ \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_t \mathbf{u}_t | \alpha_i \in \mathbb{C}, \ 1 \le i \le t \right\}$$
$$= \left\{ \sum_{i=1}^t \alpha_i \mathbf{u}_i \middle| \alpha_i \in \mathbb{C}, \ 1 \le i \le t \right\}$$

/	< l
L	/

Theorem SSS Span of a Set is a Subspace

Suppose V is a vector space. Given a set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_t} \subseteq V$, their span, $\langle S \rangle$, is a subspace.

Proof By Definition SS [709], the span contains linear combinations of vectors from the vector space V, so by repeated use of the closure properties, Property AC [660] and Property SC [660], $\langle S \rangle$ can be seen to be a subset of V.

We will then verify the three conditions of Theorem TSS [698]. First,

 $\mathbf{0} = \mathbf{0} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0}$ $= 0\mathbf{u}_1 + 0\mathbf{u}_2 + 0\mathbf{u}_3 + \dots + 0\mathbf{u}_t$ Property Z [660] for V Theorem ZSSM [676]

So we have written **0** as a linear combination of the vectors in S and by Definition SS [709], $\mathbf{0} \in \langle S \rangle$ and therefore $\langle S \rangle \neq \emptyset$.

Second, suppose $\mathbf{x} \in \langle S \rangle$ and $\mathbf{y} \in \langle S \rangle$. Can we conclude that $\mathbf{x} + \mathbf{y} \in \langle S \rangle$? What do we know about \mathbf{x} and \mathbf{y} by virtue of their membership in $\langle S \rangle$? There must be scalars from \mathbb{C} , $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_t$ and $\beta_1, \beta_2, \beta_3, \ldots, \beta_t$ so that

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_t \mathbf{u}_t$$
$$\mathbf{y} = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3 + \dots + \beta_t \mathbf{u}_t$$

Then

$$\mathbf{x} + \mathbf{y} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_t \mathbf{u}_t + \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \beta_3 \mathbf{u}_3 + \dots + \beta_t \mathbf{u}_t = \alpha_1 \mathbf{u}_1 + \beta_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \beta_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \beta_3 \mathbf{u}_3 + \dots + \alpha_t \mathbf{u}_t + \beta_t \mathbf{u}_t$$
Property AA [660], Property C [660]
$$= (\alpha_1 + \beta_1) \mathbf{u}_1 + (\alpha_2 + \beta_2) \mathbf{u}_2 + (\alpha_3 + \beta_3) \mathbf{u}_3 + \dots + (\alpha_t + \beta_t) \mathbf{u}_t$$
Property DSA [660]

Since each $\alpha_i + \beta_i$ is again a scalar from \mathbb{C} we have expressed the vector sum $\mathbf{x} + \mathbf{y}$ as a linear combination of the vectors from S, and therefore by Definition SS [709] we can say that $\mathbf{x} + \mathbf{y} \in \langle S \rangle$.

Third, suppose $\alpha \in \mathbb{C}$ and $\mathbf{x} \in \langle S \rangle$. Can we conclude that $\alpha \mathbf{x} \in \langle S \rangle$? What do we know about \mathbf{x} by virtue of its membership in $\langle S \rangle$? There must be scalars from \mathbb{C} , α_1 , α_2 , α_3 , ..., α_t so that

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_t \mathbf{u}_t$$

Then

$$\alpha \mathbf{x} = \alpha \left(\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_t \mathbf{u}_t \right)$$

= $\alpha (\alpha_1 \mathbf{u}_1) + \alpha (\alpha_2 \mathbf{u}_2) + \alpha (\alpha_3 \mathbf{u}_3) + \dots + \alpha (\alpha_t \mathbf{u}_t)$ Property DVA [660]
= $(\alpha \alpha_1) \mathbf{u}_1 + (\alpha \alpha_2) \mathbf{u}_2 + (\alpha \alpha_3) \mathbf{u}_3 + \dots + (\alpha \alpha_t) \mathbf{u}_t$ Property SMA [660]

Since each $\alpha \alpha_i$ is again a scalar from \mathbb{C} we have expressed the scalar multiple $\alpha \mathbf{x}$ as a linear combination of the vectors from S, and therefore by Definition SS [709] we can say that $\alpha \mathbf{x} \in \langle S \rangle$.

With the three conditions of Theorem TSS [698] met, we can say that $\langle S \rangle$ is a subspace (and so is also vector space, Definition VS [659]). (See Exercise SS.T20 [300], Exercise SS.T21 [300], Exercise SS.T22 [300].)

Example SSP

Span of a set of polynomials

In Example SP4 [700] we proved that

$$W = \{ p(x) | p \in P_4, \ p(2) = 0 \}$$

is a subspace of P_4 , the vector space of polynomials of degree at most 4. Since W is a vector space itself, let's construct a span within W. First let

$$S = \left\{ x^4 - 4x^3 + 5x^2 - x - 2, \ 2x^4 - 3x^3 - 6x^2 + 6x + 4 \right\}$$

and verify that S is a subset of W by checking that each of these two polynomials has x = 2 as a root. Now, if we define $U = \langle S \rangle$, then Theorem SSS [709] tells us that U is a subspace of W. So quite quickly we have built a chain of subspaces, U inside W, and W inside P_4 .

Rather than dwell on how quickly we can build subspaces, let's try to gain a better understanding of just how the span construction creates subspaces, in the context of this example. We can quickly build representative elements of U,

$$3(x^4 - 4x^3 + 5x^2 - x - 2) + 5(2x^4 - 3x^3 - 6x^2 + 6x + 4) = 13x^4 - 27x^3 - 15x^2 + 27x + 14x^4 - 27x^4 - 15x^2 + 27x + 14x^4 - 14x^4 -$$

and

$$(-2)(x^4 - 4x^3 + 5x^2 - x - 2) + 8(2x^4 - 3x^3 - 6x^2 + 6x + 4) = 14x^4 - 16x^3 - 58x^2 + 50x + 36x^2 + 50x^2 + 50x$$

and each of these polynomials must be in W since it is closed under addition and scalar multiplication. But you might check for yourself that both of these polynomials have x = 2 as a root.

I can tell you that $\mathbf{y} = 3x^4 - 7x^3 - x^2 + 7x - 2$ is not in U, but would you believe me? A first check shows that \mathbf{y} does have x = 2 as a root, but that only shows that $\mathbf{y} \in W$. What does \mathbf{y} have to do to gain membership in $U = \langle S \rangle$? It must be a linear combination of the vectors in S, $x^4 - 4x^3 + 5x^2 - x - 2$ and $2x^4 - 3x^3 - 6x^2 + 6x + 4$. So let's suppose that \mathbf{y} is such a linear combination,

$$\mathbf{y} = 3x^4 - 7x^3 - x^2 + 7x - 2$$

= $\alpha_1(x^4 - 4x^3 + 5x^2 - x - 2) + \alpha_2(2x^4 - 3x^3 - 6x^2 + 6x + 4)$
= $(\alpha_1 + 2\alpha_2)x^4 + (-4\alpha_1 - 3\alpha_2)x^3 + (5\alpha_1 - 6\alpha_2)x^2 + (-\alpha_1 + 6\alpha_2)x - (-2\alpha_1 + 4\alpha_2)$

Notice that operations above are done in accordance with the definition of the vector space of polynomials (Example VSP [665]). Now, if we equate coefficients, which is the definition of equality for polynomials, then we obtain the system of five linear equations in two variables

$$\alpha_1 + 2\alpha_2 = 3$$

$$-4\alpha_1 - 3\alpha_2 = -7$$

$$5\alpha_1 - 6\alpha_2 = -1$$

$$-\alpha_1 + 6\alpha_2 = 7$$

$$-2\alpha_1 + 4\alpha_2 = -2$$

Build an augmented matrix from the system and row-reduce,

$$\begin{bmatrix} 1 & 2 & 3 \\ -4 & -3 & -7 \\ 5 & -6 & -1 \\ -1 & 6 & 7 \\ -2 & 4 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

With a leading 1 in the final column of the row-reduced augmented matrix, Theorem RCLS [120] tells us the system of equations is inconsistent. Therefore, there are no scalars, α_1 and α_2 , to establish \mathbf{y} as a linear combination of the elements in U. So $\mathbf{y} \notin U$.

Let's again examine membership in a span.

Example SM32

A subspace of M_{32}

The set of all 3×2 matrices forms a vector space when we use the operations of matrix addition (Definition MA [427]) and scalar matrix multiplication (Definition MSM [428]), as was show in Example VSM [663]. Consider the subset

$$S = \left\{ \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -5 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 14 & -1 \end{bmatrix}, \begin{bmatrix} 3 & -1 \\ -1 & 2 \\ -19 & -11 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 1 & -2 \\ 14 & -2 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ -4 & 0 \\ -17 & 7 \end{bmatrix} \right\}$$

and define a new subset of vectors W in M_{32} using the span (Definition SS [709]), $W = \langle S \rangle$. So by Theorem SSS [709] we know that W is a subspace of M_{32} . While W is an infinite set, and this is a precise description, it would still be worthwhile to investigate whether or not W contains certain elements.

First, is

$$\mathbf{y} = \begin{bmatrix} 9 & 3\\ 7 & 3\\ 10 & -11 \end{bmatrix}$$

in W? To answer this, we want to determine if **y** can be written as a linear combination of the five matrices in S. Can we find scalars, α_1 , α_2 , α_3 , α_4 , α_5 so that

$$\begin{bmatrix} 9 & 3\\ 7 & 3\\ 10 & -11 \end{bmatrix} = \alpha_1 \begin{bmatrix} 3 & 1\\ 4 & 2\\ 5 & -5 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 & 1\\ 2 & -1\\ 14 & -1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 3 & -1\\ -1 & 2\\ -19 & -11 \end{bmatrix} + \alpha_4 \begin{bmatrix} 4 & 2\\ 1 & -2\\ 14 & -2 \end{bmatrix} + \alpha_5 \begin{bmatrix} 3 & 1\\ -4 & 0\\ -17 & 7 \end{bmatrix}$$
$$= \begin{bmatrix} 3\alpha_1 + \alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 & \alpha_1 + \alpha_2 - \alpha_3 + 2\alpha_4 + \alpha_5\\ 4\alpha_1 + 2\alpha_2 - \alpha_3 + \alpha_4 - 4\alpha_5 & 2\alpha_1 - \alpha_2 + 2\alpha_3 - 2\alpha_4\\ 5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 & -5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 \end{bmatrix}$$

Using our definition of matrix equality (Definition ME [427]) we can translate this statement into six equations in the five unknowns,

$$3\alpha_{1} + \alpha_{2} + 3\alpha_{3} + 4\alpha_{4} + 3\alpha_{5} = 9$$

$$\alpha_{1} + \alpha_{2} - \alpha_{3} + 2\alpha_{4} + \alpha_{5} = 3$$

$$4\alpha_{1} + 2\alpha_{2} - \alpha_{3} + \alpha_{4} - 4\alpha_{5} = 7$$

$$2\alpha_{1} - \alpha_{2} + 2\alpha_{3} - 2\alpha_{4} = 3$$

$$5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 = 10$$

$$-5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 = -11$$

This is a linear system of equations, which we can represent with an augmented matrix and row-reduce in search of solutions. The matrix that is row-equivalent to the augmented matrix is

So we recognize that the system is consistent since there is no leading 1 in the final column (Theorem RCLS [120]), and compute n - r = 5 - 4 = 1 free variables (Theorem FVCS [124]). While there are infinitely many solutions, we are only in pursuit of a single solution, so let's choose the free variable $\alpha_5 = 0$ for simplicity's sake. Then we easily see that $\alpha_1 = 2$, $\alpha_2 = -1$, $\alpha_3 = 0$, $\alpha_4 = 1$. So the scalars $\alpha_1 = 2$, $\alpha_2 = -1$, $\alpha_3 = 0$, $\alpha_4 = 1$, $\alpha_5 = 0$ will provide a linear combination of the elements of S that equals y, as we can verify by checking,

$$\begin{bmatrix} 9 & 3 \\ 7 & 3 \\ 10 & -11 \end{bmatrix} = 2 \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -5 \end{bmatrix} + (-1) \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 14 & -1 \end{bmatrix} + (1) \begin{bmatrix} 4 & 2 \\ 1 & -2 \\ 14 & -2 \end{bmatrix}$$

So with one particular linear combination in hand, we are convinced that \mathbf{y} deserves to be a member of $W = \langle S \rangle$. Second, is

$$\mathbf{x} = \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & -2 \end{bmatrix}$$

in W? To answer this, we want to determine if **x** can be written as a linear combination of the five matrices in S. Can we find scalars, α_1 , α_2 , α_3 , α_4 , α_5 so that

$$\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & -2 \end{bmatrix} = \alpha_1 \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -5 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 14 & -1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 3 & -1 \\ -1 & 2 \\ -19 & -11 \end{bmatrix} + \alpha_4 \begin{bmatrix} 4 & 2 \\ 1 & -2 \\ 14 & -2 \end{bmatrix} + \alpha_5 \begin{bmatrix} 3 & 1 \\ -4 & 0 \\ -17 & 7 \end{bmatrix}$$
$$= \begin{bmatrix} 3\alpha_1 + \alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 & \alpha_1 + \alpha_2 - \alpha_3 + 2\alpha_4 + \alpha_5 \\ 4\alpha_1 + 2\alpha_2 - \alpha_3 + \alpha_4 - 4\alpha_5 & 2\alpha_1 - \alpha_2 + 2\alpha_3 - 2\alpha_4 \\ 5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 & -5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 \end{bmatrix}$$

Using our definition of matrix equality (Definition ME [427]) we can translate this statement into six equations in the five unknowns,

$$3\alpha_{1} + \alpha_{2} + 3\alpha_{3} + 4\alpha_{4} + 3\alpha_{5} = 2$$

$$\alpha_{1} + \alpha_{2} - \alpha_{3} + 2\alpha_{4} + \alpha_{5} = 1$$

$$4\alpha_{1} + 2\alpha_{2} - \alpha_{3} + \alpha_{4} - 4\alpha_{5} = 3$$

$$2\alpha_{1} - \alpha_{2} + 2\alpha_{3} - 2\alpha_{4} = 1$$

$$5\alpha_{1} + 14\alpha_{2} - 19\alpha_{3} + 14\alpha_{4} - 17\alpha_{5} = 4$$

$$-5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 = -2$$

This is a linear system of equations, which we can represent with an augmented matrix and row-reduce in search of solutions. The matrix that is row-equivalent to the augmented matrix is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \frac{5}{8} & 0 \\ 0 & 1 & 0 & 0 & -\frac{38}{8} & 0 \\ 0 & 0 & 1 & 0 & -\frac{7}{8} & 0 \\ 0 & 0 & 0 & 1 & -\frac{17}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

With a leading 1 in the last column Theorem RCLS [120] tells us that the system is inconsistent. Therefore, there are no values for the scalars that will place \mathbf{x} in W, and so we conclude that $\mathbf{x} \notin W$.

Notice how Example SSP [712] and Example SM32 [714] contained questions about membership in a span, but these questions quickly became questions about solutions to a system of linear equations. This will be a common theme going forward.

Subsection SC Subspace Constructions

Several of the subsets of vectors spaces that we worked with in Chapter M [425] are also subspaces — they are closed under vector addition and scalar multiplication in \mathbb{C}^m .

Theorem CSMS Column Space of a Matrix is a Subspace

Suppose that A is an $m \times n$ matrix. Then $\mathcal{C}(A)$ is a subspace of \mathbb{C}^m .

Proof Definition CSM [560] shows us that $\mathcal{C}(A)$ is a subset of \mathbb{C}^m , and that it is defined as the span of a set of vectors from \mathbb{C}^m (the columns of the matrix). Since $\mathcal{C}(A)$ is a span, Theorem SSS [709] says it is a subspace.

That was easy! Notice that we could have used this same approach to prove that the null space is a subspace, since Theorem SSNS [283] provided a description of the null space of a matrix as the span of a set of vectors. However, I much prefer the current proof of Theorem NSMS [705]. Speaking of easy, here is a very easy theorem that exposes another of our constructions as creating subspaces.

Theorem RSMS

Row Space of a Matrix is a Subspace

Suppose that A is an $m \times n$ matrix. Then $\mathcal{R}(A)$ is a subspace of \mathbb{C}^n .

Proof Definition RSM [577] says $\mathcal{R}(A) = \mathcal{C}(A^t)$, so the row space of a matrix is a column

space, and every column space is a subspace by Theorem CSMS [719]. That's enough.

One more.

Theorem LNSMS Left Null Space of a Matrix is a Subspace

Suppose that A is an $m \times n$ matrix. Then $\mathcal{L}(A)$ is a subspace of \mathbb{C}^m .

Proof Definition LNS [609] says $\mathcal{L}(A) = \mathcal{N}(A^t)$, so the left null space is a null space, and every null space is a subspace by Theorem NSMS [705]. Done.

So the span of a set of vectors, and the null space, column space, row space and left null space of a matrix are all subspaces, and hence are all vector spaces, meaning they have all the properties detailed in Definition VS [659] and in the basic theorems presented in Section VS [659]. We have worked with these objects as just sets in Chapter V [197] and Chapter M [425], but now we understand that they have much more structure. In particular, being closed under vector addition and scalar multiplication means a subspace is also closed under linear combinations.

Subsection READ Reading Questions

1. Summarize the three conditions that allow us to quickly test if a set is a subspace.

2. Consider the set of vectors

$$W = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \middle| 3a - 2b + c = 5 \right\}$$

Is the set W a subspace of \mathbb{C}^3 ? Explain your answer.

3. Name five general constructions of sets of column vectors (subsets of \mathbb{C}^m) that we now know as subspaces.

Subsection EXC Exercises

Working within the vector space \mathbb{C}^3 , determine if $\mathbf{b} = \begin{bmatrix} 4\\ 3\\ 1 \end{bmatrix}$ is in the subspace W, C15

$$W = \left\langle \left\{ \begin{bmatrix} 3\\2\\3 \end{bmatrix}, \begin{bmatrix} 1\\0\\3 \end{bmatrix}, \begin{bmatrix} 1\\1\\0 \end{bmatrix}, \begin{bmatrix} 2\\1\\3 \end{bmatrix} \right\} \right\rangle$$

Contributed by Chris Black Solution [726]

Working within the vector space \mathbb{C}^4 , determine if $\mathbf{b} = \begin{bmatrix} 1\\1\\0\\1 \end{bmatrix}$ is in the subspace W, C16

$$W = \left\langle \left\{ \begin{bmatrix} 1\\2\\-1\\1 \end{bmatrix}, \begin{bmatrix} 1\\0\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\1\\1\\2 \end{bmatrix} \right\} \right\rangle$$

Contributed by Chris Black Solution [726]

C17 Working within the vector space \mathbb{C}^4 , determine if $\mathbf{b} = \begin{bmatrix} 2\\1\\2\\1 \end{bmatrix}$ is in the subspace W,

$$W = \left\langle \left\{ \begin{bmatrix} 1\\2\\0\\2 \end{bmatrix}, \begin{bmatrix} 1\\0\\3\\1 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\2 \end{bmatrix}, \begin{bmatrix} 1\\1\\2\\0 \end{bmatrix} \right\} \right\rangle$$

Contributed by Chris Black Solution [727]

C20 Working within the vector space P_3 of polynomials of degree 3 or less, determine if $p(x) = x^3 + 6x + 4$ is in the subspace W below.

$$W = \left\langle \left\{ x^3 + x^2 + x, \, x^3 + 2x - 6, \, x^2 - 5 \right\} \right\rangle$$

Contributed by Robert Beezer Solution [727]

C21 Consider the subspace

$$W = \left\langle \left\{ \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} -3 & 1 \\ 2 & 1 \end{bmatrix} \right\} \right\rangle$$

of the vector space of 2×2 matrices, M_{22} . Is $C = \begin{bmatrix} -3 & 3 \\ 6 & -4 \end{bmatrix}$ an element of W? Contributed by Robert Beezer Solution [729]

C25 Show that the set $W = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| 3x_1 - 5x_2 = 12 \right\}$ from Example NSC2Z [702] fails Property AC [660] and Property SC [660]. Contributed by Robert Beezer

C26 Show that the set $Y = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle| x_1 \in \mathbb{Z}, x_2 \in \mathbb{Z} \right\}$ from Example NSC2S [703] has Property AC [660]. Contributed by Robert Beezer

M20 In \mathbb{C}^3 , the vector space of column vectors of size 3, prove that the set Z is a subspace.

$$Z = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \middle| 4x_1 - x_2 + 5x_3 = 0 \right\}$$

Contributed by Robert Beezer Solution [730]

T20 A square matrix A of size n is upper triangular if $[A]_{ij} = 0$ whenever i > j. Let UT_n be the set of all upper triangular matrices of size n. Prove that UT_n is a subspace of the vector space of all square matrices of size n, M_{nn} . Contributed by Robert Beezer Solution [732] **T30** Let P be the set of all polynomials, of any degree. The set P is a vector space. Let E be the subset of P consisting of all polynomials with only terms of even degree. Prove or disprove: the set E is a subspace of P. Contributed by Chris Black Solution [733]

T31 Let P be the set of all polynomials, of any degree. The set P is a vector space. Let F be the subset of P consisting of all polynomials with only terms of odd degree. Prove or disprove: the set F is a subspace of P.

Contributed by Chris Black Solution [733]

Subsection SOL Solutions

C15 Contributed by Chris Black Statement [722] For **b** to be an element of $W = \langle S \rangle$ there must be linear combination of the vectors in S that equals **b** (Definition SSCV [270]). The existence of such scalars is equivalent to the linear system $\mathcal{LS}(A, \mathbf{b})$ being consistent, where A is the matrix whose columns are the vectors from S (Theorem SLSLC [229]).

$$\begin{bmatrix} 3 & 1 & 1 & 2 & 4 \\ 2 & 0 & 1 & 1 & 3 \\ 3 & 3 & 0 & 3 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1/2 & 1/2 & 0 \\ 0 & 1 & -1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

So by Theorem RCLS [120] the system is inconsistent, which indicates that **b** is not an element of the subspace W.

C16 Contributed by Chris Black Statement [722] For **b** to be an element of $W = \langle S \rangle$ there must be linear combination of the vectors in S that equals **b** (Definition SSCV [270]). The existence of such scalars is equivalent to the linear system $\mathcal{LS}(A, \mathbf{b})$ being consistent, where A is the matrix whose columns are the vectors from S (Theorem SLSLC [229]).

[1]	1	2	1]		1	0	0	1/3
2	0	1	1	$\xrightarrow{\text{RREF}}$	0	1	0	0
-1	3	1	0	$\rightarrow\rightarrow$	0	0	1	1/3
1	1	2	1		0	0	$\overline{0}$	$\begin{bmatrix} 1/0\\0 \end{bmatrix}$

So by Theorem RCLS [120] the system is consistent, which indicates that **b** is in the subspace W.

C17 Contributed by Chris Black Statement [723]

For **b** to be an element of $W = \langle S \rangle$ there must be linear combination of the vectors in S that equals **b** (Definition SSCV [270]). The existence of such scalars is equivalent to the linear system $\mathcal{LS}(A, \mathbf{b})$ being consistent, where A is the matrix whose columns are the vectors from S (Theorem SLSLC [229]).

[1	1	0	1	2		$\left[1 \right]$	0	0	0	3/2]
2	0	1	1	1	$\xrightarrow{\text{RREF}}$	0	1	0	0	1
0	3	0	2	2	\rightarrow	0	0	1	0	-3/2
2	1	2	0	0		0	0	0	1	-1/2

So by Theorem RCLS [120] the system is consistent, which indicates that **b** is in the subspace W.

C20 Contributed by Robert Beezer Statement [723] The question is if p can be written as a linear combination of the vectors in W. To check this,

we set p equal to a linear combination and massage with the definitions of vector addition and scalar multiplication that we get with P_3 (Example VSP [665])

$$p(x) = a_1(x^3 + x^2 + x) + a_2(x^3 + 2x - 6) + a_3(x^2 - 5)$$

$$x^3 + 6x + 4 = (a_1 + a_2)x^3 + (a_1 + a_3)x^2 + (a_1 + 2a_2)x + (-6a_2 - 5a_3)$$

Equating coefficients of equal powers of x, we get the system of equations,

$$a_1 + a_2 = 1$$

 $a_1 + a_3 = 0$
 $a_1 + 2a_2 = 6$
 $-6a_2 - 5a_3 = 4$

The augmented matrix of this system of equations row-reduces to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

There is a leading 1 in the last column, so Theorem RCLS [120] implies that the system is inconsistent. So there is no way for p to gain membership in W, so $p \notin W$.

C21 Contributed by Robert Beezer Statement [723]

In order to belong to W, we must be able to express C as a linear combination of the elements in the spanning set of W. So we begin with such an expression, using the unknowns a, b, cfor the scalars in the linear combination.

$$C = \begin{bmatrix} -3 & 3\\ 6 & -4 \end{bmatrix} = a \begin{bmatrix} 2 & 1\\ 3 & -1 \end{bmatrix} + b \begin{bmatrix} 4 & 0\\ 2 & 3 \end{bmatrix} + c \begin{bmatrix} -3 & 1\\ 2 & 1 \end{bmatrix}$$

Massaging the right-hand side, according to the definition of the vector space operations in M_{22} (Example VSM [663]), we find the matrix equality,

$$\begin{bmatrix} -3 & 3\\ 6 & -4 \end{bmatrix} = \begin{bmatrix} 2a+4b-3c & a+c\\ 3a+2b+2c & -a+3b+c \end{bmatrix}$$

Matrix equality allows us to form a system of four equations in three variables, whose augmented matrix row-reduces as follows,

$$\begin{bmatrix} 2 & 4 & -3 & -3 \\ 1 & 0 & 1 & 3 \\ 3 & 2 & 2 & 6 \\ -1 & 3 & 1 & -4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since this system of equations is consistent (Theorem RCLS [120]), a solution will provide values for a, b and c that allow us to recognize C as an element of W.

M20 Contributed by Robert Beezer Statement [724]

The membership criteria for Z is a single linear equation, which comprises a homogeneous system of equations. As such, we can recognize Z as the solutions to this system, and therefore Z is a null space. Specifically, $Z = \mathcal{N}(\begin{bmatrix} 4 & -1 & 5 \end{bmatrix})$. Every null space is a subspace by Theorem NSMS [705].

A less direct solution appeals to Theorem TSS [698].

First, we want to be certain Z is non-empty. The zero vector of \mathbb{C}^3 , $\mathbf{0} = \begin{bmatrix} 0\\0\\0 \end{bmatrix}$, is a good candidate, since if it fails to be in Z, we will know that Z is *not* a vector space. Check that

4(0) - (0) + 5(0) = 0

so that $\mathbf{0} \in \mathbb{Z}$. Suppose $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$ are vectors from \mathbb{Z} . Then we know that these vectors cannot be totally arbitrary, they must have gained membership in \mathbb{Z} by virtue of meeting the membership test. For example, we know that \mathbf{x} must satisfy $4x_1 - x_2 + 5x_3 = 0$ while \mathbf{y} must satisfy $4y_1 - y_2 + 5y_3 = 0$. Our second criteria asks the question, is $\mathbf{x} + \mathbf{y} \in \mathbb{Z}$? Notice first that

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix}$$

and we can test this vector for membership in Z as follows,

$$\begin{aligned} 4(x_1 + y_1) &- 1(x_2 + y_2) + 5(x_3 + y_3) \\ &= 4x_1 + 4y_1 - x_2 - y_2 + 5x_3 + 5y_3 \\ &= (4x_1 - x_2 + 5x_3) + (4y_1 - y_2 + 5y_3) \\ &= 0 + 0 \qquad \qquad \mathbf{x} \in \mathbb{Z}, \ \mathbf{y} \in \mathbb{Z} \\ &= 0 \end{aligned}$$

and by this computation we see that $\mathbf{x} + \mathbf{y} \in Z$.

If α is a scalar and $\mathbf{x} \in \mathbb{Z}$, is it always true that $\alpha \mathbf{x} \in \mathbb{Z}$? To check our third criteria, we examine

$$\alpha \mathbf{x} = \alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{bmatrix}$$

and we can test this vector for membership in Z with

$$4(\alpha x_1) - (\alpha x_2) + 5(\alpha x_3)$$

= $\alpha (4x_1 - x_2 + 5x_3)$
= $\alpha 0$
= 0
x $\in Z$

and we see that indeed $\alpha \mathbf{x} \in Z$. With the three conditions of Theorem TSS [698] fulfilled, we can conclude that Z is a subspace of \mathbb{C}^3 .

T20 Contributed by Robert Beezer Statement [724] Apply Theorem TSS [698].

First, the zero vector of M_{nn} is the zero matrix, \mathcal{O} , whose entries are all zero (Definition ZM [433]). This matrix then meets the condition that $[\mathcal{O}]_{ij} = 0$ for i > j and so is an element of UT_n .

Suppose $A, B \in UT_n$. Is $A + B \in UT_n$? We examine the entries of A + B "below" the diagonal. That is, in the following, assume that i > j.

$$[A + B]_{ij} = [A]_{ij} + [B]_{ij}$$

= 0 + 0
= 0
$$A, B \in UT_n$$

which qualifies A + B for membership in UT_n .

Suppose $\alpha \in \mathbb{C}$ and $A \in UT_n$. Is $\alpha A \in UT_n$? We examine the entries of αA "below" the diagonal. That is, in the following, assume that i > j.

 $\begin{aligned} [\alpha A]_{ij} &= \alpha \, [A]_{ij} & \text{Definition MSM [428]} \\ &= \alpha 0 & A \in UT_n \\ &= 0 & \end{aligned}$

which qualifies αA for membership in UT_n .

Having fulfilled the three conditions of Theorem TSS [698] we see that UT_n is a subspace of M_{nn} .

T30 Contributed by Chris Black Statement [725]

Proof: Let *E* be the subset of *P* comprised of all polynomials with all terms of even degree. Clearly the set *E* is non-empty, as z(x) = 0 is a polynomial of even degree. Let p(x) and q(x) be arbitrary elements of *E*. Then there exist nonnegative integers *m* and *n* so that

$$p(x) = a_0 + a_2 x^2 + a_4 x^4 + \dots + a_{2n} x^{2n}$$

$$q(x) = b_0 + b_2 x^2 + b_4 x^4 + \dots + b_{2m} x^{2m}$$

for some constants a_0, a_2, \ldots, a_{2n} and b_0, b_2, \ldots, b_{2m} . Without loss of generality, we can assume that $m \leq n$. Thus, we have

$$p(x) + q(x) = (a_0 + b_0) + (a_2 + b_2)x^2 + \dots + (a_{2m} + b_{2m})x^{2m} + a_{2m+2}x^{2m+2} + \dots + a_{2n}x^{2n}$$

so p(x) + q(x) has all even terms, and thus $p(x) + q(x) \in E$. Similarly, let α be a scalar. Then

$$\alpha p(x) = \alpha (a_0 + a_2 x^2 + a_4 x^4 + \dots + a_{2n} x^{2n})$$

= $\alpha a_0 + (\alpha a_2) x^2 + (\alpha a_4) x^4 + \dots + (\alpha a_{2n}) x^{2n}$

so that $\alpha p(x)$ also has only terms of even degree, and $\alpha p(x) \in E$. Thus, E is a subspace of P.

T31 Contributed by Chris Black Statement [725]

This conjecture is false. We know that the zero vector in P is the polynomial z(x) = 0, which does not have odd degree. Thus, the set F does not contain the zero vector, and cannot be a vector space.

Section LISS Linear Independence and Spanning Sets

A vector space is defined as a set with two operations, meeting ten properties (Definition VS [659]). Just as the definition of span of a set of vectors only required knowing how to add vectors and how to multiply vectors by scalars, so it is with linear independence. A definition of a linear independent set of vectors in an arbitrary vector space only requires knowing how to form linear combinations and equating these with the zero vector. Since every vector space must have a zero vector (Property Z [660]), we always have a zero vector at our disposal.

In this section we will also put a twist on the notion of the span of a set of vectors. Rather than beginning with a set of vectors and creating a subspace that is the span, we will instead begin with a subspace and look for a set of vectors whose span equals the subspace.

The combination of linear independence and spanning will be very important going forward.

Subsection LI Linear Independence

Our previous definition of linear independence (Definition LICV [317]) employed a relation of linear dependence that was a linear combination on one side of an equality and a zero vector on the other side. As a linear combination in a vector space (Definition LC [707]) depends only on vector addition and scalar multiplication, and every vector space must have a zero vector (Property Z [660]), we can extend our definition of linear independence from the setting of \mathbb{C}^m to the setting of a general vector space V with almost no changes. Compare these next two definitions with Definition RLDCV [316] and Definition LICV [317].

Definition RLD Relation of Linear Dependence

Suppose that V is a vector space. Given a set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$, an equation of the form

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 + \dots + \alpha_n \mathbf{u}_n = \mathbf{0}$$

is a **relation of linear dependence** on S. If this equation is formed in a trivial fashion, i.e. $\alpha_i = 0, 1 \leq i \leq n$, then we say it is a **trivial relation of linear dependence** on S. \triangle

Definition LI Linear Independence

Suppose that V is a vector space. The set of vectors $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ from V is

linearly dependent if there is a relation of linear dependence on S that is not trivial. In the case where the *only* relation of linear dependence on S is the trivial one, then S is a **linearly independent** set of vectors. \triangle

Notice the emphasis on the word "only." This might remind you of the definition of a nonsingular matrix, where if the matrix is employed as the coefficient matrix of a homogeneous system then the *only* solution is the *trivial* one.

Example LIP4 Linear independence in P_4

In the vector space of polynomials with degree 4 or less, P_4 (Example VSP [665]) consider the set

$$S = \left\{ 2x^4 + 3x^3 + 2x^2 - x + 10, \ -x^4 - 2x^3 + x^2 + 5x - 8, \ 2x^4 + x^3 + 10x^2 + 17x - 2 \right\}.$$

Is this set of vectors linearly independent or dependent? Consider that

$$3 (2x^{4} + 3x^{3} + 2x^{2} - x + 10) + 4 (-x^{4} - 2x^{3} + x^{2} + 5x - 8) + (-1) (2x^{4} + x^{3} + 10x^{2} + 17x - 2) = 0x^{4} + 0x^{3} + 0x^{2} + 0x + 0 = 0$$

This is a nontrivial relation of linear dependence (Definition RLD [735]) on the set S and so convinces us that S is linearly dependent (Definition LI [735]).

Now, I hear you say, "Where did *those* scalars come from?" Do not worry about that right now, just be sure you understand why the above explanation is sufficient to prove that S is linearly dependent. The remainder of the example will demonstrate how we might find

these scalars if they had not been provided so readily. Let's look at another set of vectors (polynomials) from P_4 . Let

$$T = \left\{ 3x^4 - 2x^3 + 4x^2 + 6x - 1, \ -3x^4 + 1x^3 + 0x^2 + 4x + 2, \\ 4x^4 + 5x^3 - 2x^2 + 3x + 1, \ 2x^4 - 7x^3 + 4x^2 + 2x + 1 \right\}$$

Suppose we have a relation of linear dependence on this set,

$$0 = 0x^{4} + 0x^{3} + 0x^{2} + 0x + 0$$

= $\alpha_{1} (3x^{4} - 2x^{3} + 4x^{2} + 6x - 1) + \alpha_{2} (-3x^{4} + 1x^{3} + 0x^{2} + 4x + 2)$
+ $\alpha_{3} (4x^{4} + 5x^{3} - 2x^{2} + 3x + 1) + \alpha_{4} (2x^{4} - 7x^{3} + 4x^{2} + 2x + 1)$

Using our definitions of vector addition and scalar multiplication in P_4 (Example VSP [665]), we arrive at,

$$0x^{4} + 0x^{3} + 0x^{2} + 0x + 0 = (3\alpha_{1} - 3\alpha_{2} + 4\alpha_{3} + 2\alpha_{4})x^{4} + (-2\alpha_{1} + \alpha_{2} + 5\alpha_{3} - 7\alpha_{4})x^{3} + (4\alpha_{1} + -2\alpha_{3} + 4\alpha_{4})x^{2} + (6\alpha_{1} + 4\alpha_{2} + 3\alpha_{3} + 2\alpha_{4})x + (-\alpha_{1} + 2\alpha_{2} + \alpha_{3} + \alpha_{4}).$$

Equating coefficients, we arrive at the homogeneous system of equations,

$$3\alpha_1 - 3\alpha_2 + 4\alpha_3 + 2\alpha_4 = 0 -2\alpha_1 + \alpha_2 + 5\alpha_3 - 7\alpha_4 = 0$$

$$4\alpha_{1} + -2\alpha_{3} + 4\alpha_{4} = 0$$

$$6\alpha_{1} + 4\alpha_{2} + 3\alpha_{3} + 2\alpha_{4} = 0$$

$$-\alpha_{1} + 2\alpha_{2} + \alpha_{3} + \alpha_{4} = 0$$

We form the coefficient matrix of this homogeneous system of equations and row-reduce to find

$\lceil 1 \rceil$	0	0	0]
0	1	0	0
0	0	1	0
0	0	0	1
0	0	0	0

We expected the system to be consistent (Theorem HSC [147]) and so can compute n - r = 4 - 4 = 0 and Theorem CSRN [123] tells us that the solution is unique. Since this is a homogeneous system, this unique solution is the trivial solution (Definition TSHSE [148]), $\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 0, \alpha_4 = 0$. So by Definition LI [735] the set T is linearly independent.

A few observations. If we had discovered infinitely many solutions, then we could have used one of the non-trivial ones to provide a linear combination in the manner we used to show that S was linearly dependent. It is important to realize that it is not interesting that we can create a relation of linear dependence with zero scalars — we can *always* do that — but that for T, this is the *only* way to create a relation of linear dependence. It was no accident that we arrived at a homogeneous system of equations in this example, it is related to our use of the zero vector in defining a relation of linear dependence. It is easy to present a convincing statement that a set is linearly dependent (just exhibit a nontrivial relation of linear dependence) but a convincing statement of linear independence requires demonstrating that there is no relation of linear dependence other than the trivial one. Notice how we relied on theorems from Chapter SLE [2] to provide this demonstration. Whew! There's a lot going on in this example. Spend some time with it, we'll be waiting patiently right here when you get back.

Example LIM32

Linear independence in M_{32}

Consider the two sets of vectors R and S from the vector space of all 3×2 matrices, M_{32} (Example VSM [663])

$$R = \left\{ \begin{bmatrix} 3 & -1 \\ 1 & 4 \\ 6 & -6 \end{bmatrix}, \begin{bmatrix} -2 & 3 \\ 1 & -3 \\ -2 & -6 \end{bmatrix}, \begin{bmatrix} 6 & -6 \\ -1 & 0 \\ 7 & -9 \end{bmatrix}, \begin{bmatrix} 7 & 9 \\ -4 & -5 \\ 2 & 5 \end{bmatrix} \right\}$$
$$S = \left\{ \begin{bmatrix} 2 & 0 \\ 1 & -1 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} -4 & 0 \\ -2 & 2 \\ -2 & -6 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} -5 & 3 \\ -10 & 7 \\ 2 & 0 \end{bmatrix} \right\}$$

One set is linearly independent, the other is not. Which is which? Let's examine R first. Build a generic relation of linear dependence (Definition RLD [735]),

$$\alpha_1 \begin{bmatrix} 3 & -1 \\ 1 & 4 \\ 6 & -6 \end{bmatrix} + \alpha_2 \begin{bmatrix} -2 & 3 \\ 1 & -3 \\ -2 & -6 \end{bmatrix} + \alpha_3 \begin{bmatrix} 6 & -6 \\ -1 & 0 \\ 7 & -9 \end{bmatrix} + \alpha_4 \begin{bmatrix} 7 & 9 \\ -4 & -5 \\ 2 & 5 \end{bmatrix} = \mathbf{0}$$

Massaging the left-hand side with our definitions of vector addition and scalar multiplication in M_{32} (Example VSM [663]) we obtain,

$$\begin{bmatrix} 3\alpha_1 - 2\alpha_2 + 6\alpha_3 + 7\alpha_4 & -1\alpha_1 + 3\alpha_2 - 6\alpha_3 + 9\alpha_4 \\ 1\alpha_1 + 1\alpha_2 - \alpha_3 - 4\alpha_4 & 4\alpha_1 - 3\alpha_2 + -5\alpha_4 \\ 6\alpha_1 - 2\alpha_2 + 7\alpha_3 + 2\alpha_4 & -6\alpha_1 - 6\alpha_2 - 9\alpha_3 + 5\alpha_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Using our definition of matrix equality (Definition ME [427]) and equating corresponding entries we get the homogeneous system of six equations in four variables,

$$3\alpha_{1} - 2\alpha_{2} + 6\alpha_{3} + 7\alpha_{4} = 0$$

$$-1\alpha_{1} + 3\alpha_{2} - 6\alpha_{3} + 9\alpha_{4} = 0$$

$$1\alpha_{1} + 1\alpha_{2} - \alpha_{3} - 4\alpha_{4} = 0$$

$$4\alpha_{1} - 3\alpha_{2} + -5\alpha_{4} = 0$$

$$6\alpha_{1} - 2\alpha_{2} + 7\alpha_{3} + 2\alpha_{4} = 0$$

$$-6\alpha_{1} - 6\alpha_{2} - 9\alpha_{3} + 5\alpha_{4} = 0$$

Form the coefficient matrix of this homogeneous system and row-reduce to obtain

[1]	0	0	0]
0	1	0	0
0	0	1	0
0	0	0	1
0	0	0	$\overline{0}$
0	0	0	0

Analyzing this matrix we are led to conclude that $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$, $\alpha_4 = 0$. This means there is *only* a trivial relation of linear dependence on the vectors of R and so we call R a linearly independent set (Definition LI [735]).

So it must be that S is linearly dependent. Let's see if we can find a non-trivial relation of linear dependence on S. We will begin as with R, by constructing a relation of linear dependence (Definition RLD [735]) with unknown scalars,

$$\alpha_1 \begin{bmatrix} 2 & 0 \\ 1 & -1 \\ 1 & 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} -4 & 0 \\ -2 & 2 \\ -2 & -6 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ 2 & 4 \end{bmatrix} + \alpha_4 \begin{bmatrix} -5 & 3 \\ -10 & 7 \\ 2 & 0 \end{bmatrix} = \mathbf{0}$$

Massaging the left-hand side with our definitions of vector addition and scalar multiplication in M_{32} (Example VSM [663]) we obtain,

$$\begin{bmatrix} 2\alpha_1 - 4\alpha_2 + \alpha_3 - 5\alpha_4 & \alpha_3 + 3\alpha_4 \\ \alpha_1 - 2\alpha_2 - 2\alpha_3 - 10\alpha_4 & -\alpha_1 + 2\alpha_2 + \alpha_3 + 7\alpha_4 \\ \alpha_1 - 2\alpha_2 + 2\alpha_3 + 2\alpha_4 & 3\alpha_1 - 6\alpha_2 + 4\alpha_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Using our definition of matrix equality (Definition ME [427]) and equating corresponding entries we get the homogeneous system of six equations in four variables,

$$2\alpha_1 - 4\alpha_2 + \alpha_3 - 5\alpha_4 = 0$$
$$+\alpha_3 + 3\alpha_4 = 0$$
$$\alpha_1 - 2\alpha_2 - 2\alpha_3 - 10\alpha_4 = 0$$

$$-\alpha_{1} + 2\alpha_{2} + \alpha_{3} + 7\alpha_{4} = 0$$

$$\alpha_{1} - 2\alpha_{2} + 2\alpha_{3} + 2\alpha_{4} = 0$$

$$3\alpha_{1} - 6\alpha_{2} + 4\alpha_{3} = 0$$

Form the coefficient matrix of this homogeneous system and row-reduce to obtain

$\left[1 \right]$	-2	0	-4
0	0	1	3
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Analyzing this we see that the system is consistent (we expected this since the system is homogeneous, Theorem HSC [147]) and has n - r = 4 - 2 = 2 free variables, namely α_2 and α_4 . This means there are infinitely many solutions, and in particular, we can find a non-trivial solution, so long as we do not pick all of our free variables to be zero. The mere presence of a nontrivial solution for these scalars is enough to conclude that S is a linearly dependent set (Definition LI [735]). But let's go ahead and explicitly construct a non-trivial relation of linear dependence.

Choose $\alpha_2 = 1$ and $\alpha_4 = -1$. There is nothing special about this choice, there are infinitely many possibilities, some "easier" than this one, just avoid picking both variables to be zero. Then we find the corresponding dependent variables to be $\alpha_1 = -2$ and $\alpha_3 = 3$.

So the relation of linear dependence,

$$(-2)\begin{bmatrix}2 & 0\\1 & -1\\1 & 3\end{bmatrix} + (1)\begin{bmatrix}-4 & 0\\-2 & 2\\-2 & -6\end{bmatrix} + (3)\begin{bmatrix}1 & 1\\-2 & 1\\2 & 4\end{bmatrix} + (-1)\begin{bmatrix}-5 & 3\\-10 & 7\\2 & 0\end{bmatrix} = \begin{bmatrix}0 & 0\\0 & 0\\0 & 0\end{bmatrix}$$

is an iron-clad demonstration that S is linearly dependent. Can you construct another such demonstration? $\hfill \boxtimes$

Example LIC Linearly independent set in the crazy vector space

Is the set $R = \{(1, 0), (6, 3)\}$ linearly independent in the crazy vector space C (Example CVS [670])? We begin with an arbitrary relation of linear independence on R

 $\mathbf{0} = a_1(1, 0) + a_2(6, 3)$ Definition RLD [735]

and then massage it to a point where we can apply the definition of equality in C. Recall the definitions of vector addition and scalar multiplication in C are not what you would expect.

$$(-1, -1) = \mathbf{0}$$

= $a_1(1, 0) + a_2(6, 3)$
= $(1a_1 + a_1 - 1, 0a_1 + a_1 - 1) + (6a_2 + a_2 - 1, 3a_2 + a_2 - 1)$
= $(2a_1 - 1, a_1 - 1) + (7a_2 - 1, 4a_2 - 1)$
Example CVS [670]
Example CVS [670]

$$= (2a_1 - 1 + 7a_2 - 1 + 1, a_1 - 1 + 4a_2 - 1 + 1)$$

= $(2a_1 + 7a_2 - 1, a_1 + 4a_2 - 1)$
Example CVS [670]

Equality in C (Example CVS [670]) then yields the two equations,

$$2a_1 + 7a_2 - 1 = -1$$
$$a_1 + 4a_2 - 1 = -1$$

which becomes the homogeneous system

$$2a_1 + 7a_2 = 0$$

$$a_1 + 4a_2 = 0$$

Since the coefficient matrix of this system is nonsingular (check this!) the system has only the trivial solution $a_1 = a_2 = 0$. By Definition LI [735] the set R is linearly independent. Notice that even though the zero vector of C is not what we might first suspected, a question about linear independence still concludes with a question about a homogeneous system of equations. Hmmm.

Subsection SS Spanning Sets

In a vector space V, suppose we are given a set of vectors $S \subseteq V$. Then we can immediately construct a subspace, $\langle S \rangle$, using Definition SS [709] and then be assured by Theorem SSS [709] that the construction does provide a subspace. We now turn the situation upsidedown. Suppose we are first given a subspace $W \subseteq V$. Can we find a set S so that $\langle S \rangle = W$? Typically W is infinite and we are searching for a finite set of vectors S that we can combine in linear combinations and "build" all of W.

I like to think of S as the raw materials that are sufficient for the construction of W. If you have nails, lumber, wire, copper pipe, drywall, plywood, carpet, shingles, paint (and a few other things), then you can combine them in many different ways to create a house (or infinitely many different houses for that matter). A fast-food restaurant may have beef, chicken, beans, cheese, tortillas, taco shells and hot sauce and from this small list of ingredients build a wide variety of items for sale. Or maybe a better analogy comes from Ben Cordes — the additive primary colors (red, green and blue) can be combined to create many different colors by varying the intensity of each. The intensity is like a scalar multiple, and the combination of the three intensities is like vector addition. The three individual colors, red, green and blue, are the elements of the spanning set.

Because we will use terms like "spanned by" and "spanning set," there is the potential for confusion with "the span." Come back and reread the first paragraph of this subsection

whenever you are uncertain about the difference. Here's the working definition.

Definition TSVS To Span a Vector Space

Suppose V is a vector space. A subset S of V is a **spanning set** for V if $\langle S \rangle = V$. In this case, we also say S **spans** V.

The definition of a spanning set requires that two sets (subspaces actually) be equal. If S is a subset of V, then $\langle S \rangle \subseteq V$, always. Thus it is usually only necessary to prove that $V \subseteq \langle S \rangle$. Now would be a good time to review Definition SE [1604].

Example SSP4 Spanning set in P_4 In Example SP4 [700] we showed that

 $W = \{ p(x) | p \in P_4, \ p(2) = 0 \}$

is a subspace of P_4 , the vector space of polynomials with degree at most 4 (Example VSP [665]). In this example, we will show that the set

$$S = \left\{ x - 2, \ x^2 - 4x + 4, \ x^3 - 6x^2 + 12x - 8, \ x^4 - 8x^3 + 24x^2 - 32x + 16 \right\}$$

is a spanning set for W. To do this, we require that $W = \langle S \rangle$. This is an equality of sets. We can check that every polynomial in S has x = 2 as a root and therefore $S \subseteq W$. Since W is closed under addition and scalar multiplication, $\langle S \rangle \subseteq W$ also. So it remains to show that $W \subseteq \langle S \rangle$ (Definition SE [1604]). To do this, begin by choosing an arbitrary polynomial in W, say $r(x) = ax^4 + bx^3 + cx^2 + dx + e \in W$. This polynomial is not as arbitrary as it would appear, since we also know it must have x = 2 as a root. This translates to

$$0 = a(2)^{4} + b(2)^{3} + c(2)^{2} + d(2) + e = 16a + 8b + 4c + 2d + e$$

as a condition on r.

We wish to show that r is a polynomial in $\langle S \rangle$, that is, we want to show that r can be written as a linear combination of the vectors (polynomials) in S. So let's try.

$$r(x) = ax^{4} + bx^{3} + cx^{2} + dx + e$$

= $\alpha_{1} (x - 2) + \alpha_{2} (x^{2} - 4x + 4) + \alpha_{3} (x^{3} - 6x^{2} + 12x - 8)$
+ $\alpha_{4} (x^{4} - 8x^{3} + 24x^{2} - 32x + 16)$
= $\alpha_{4}x^{4} + (\alpha_{3} - 8\alpha_{4}) x^{3} + (\alpha_{2} - 6\alpha_{3} + 24\alpha_{4}) x^{2}$
+ $(\alpha_{1} - 4\alpha_{2} + 12\alpha_{3} - 32\alpha_{4}) x + (-2\alpha_{1} + 4\alpha_{2} - 8\alpha_{3} + 16\alpha_{4})$

Equating coefficients (vector equality in P_4) gives the system of five equations in four variables,

$$\alpha_4 = a$$
$$\alpha_3 - 8\alpha_4 = b$$
$$\alpha_2 - 6\alpha_3 + 24\alpha_4 = c$$

$$\alpha_1 - 4\alpha_2 + 12\alpha_3 - 32\alpha_4 = d -2\alpha_1 + 4\alpha_2 - 8\alpha_3 + 16\alpha_4 = e$$

Any solution to this system of equations will provide the linear combination we need to determine if $r \in \langle S \rangle$, but we need to be convinced there is a solution for any values of a, b, c, d, e that qualify r to be a member of W. So the question is: is this system of equations consistent? We will form the augmented matrix, and row-reduce. (We probably need to do this by hand, since the matrix is symbolic — reversing the order of the first four rows is the best way to start). We obtain a matrix in reduced row-echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 32a + 12b + 4c + d \\ 0 & 1 & 0 & 0 & 24a + 6b + c \\ 0 & 0 & 1 & 0 & 8a + b \\ 0 & 0 & 0 & 1 & a \\ 0 & 0 & 0 & 0 & 16a + 8b + 4c + 2d + e \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 32a + 12b + 4c + d \\ 0 & 1 & 0 & 0 & 24a + 6b + c \\ 0 & 0 & 1 & 0 & 8a + b \\ 0 & 0 & 0 & 1 & a \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

For your results to match our first matrix, you may find it necessary to multiply the final row of your row-reduced matrix by the appropriate scalar, and/or add multiples of this row to some of the other rows. To obtain the second version of the matrix, the last entry of the last column has been simplified to zero according to the one condition we were able to impose on an arbitrary polynomial from W. So with no leading 1's in the last column, Theorem RCLS [120] tells us this system is consistent. Therefore, *any* polynomial from W can be written as

a linear combination of the polynomials in S, so $W \subseteq \langle S \rangle$. Therefore, $W = \langle S \rangle$ and S is a spanning set for W by Definition TSVS [746].

Notice that an alternative to row-reducing the augmented matrix by hand would be to appeal to Theorem FS [625] by expressing the column space of the coefficient matrix as a null space, and then verifying that the condition on r guarantees that r is in the column space, thus implying that the system is always consistent. Give it a try, we'll wait. This has been a complicated example, but worth studying carefully.

Given a subspace and a set of vectors, as in Example SSP4 [746] it can take some work to determine that the set actually is a spanning set. An even harder problem is to be confronted with a subspace and required to construct a spanning set with no guidance. We will now work an example of this flavor, but some of the steps will be unmotivated. Fortunately, we will have some better tools for this type of problem later on.

Example SSM22

Spanning set in M_{22}

In the space of all 2×2 matrices, M_{22} consider the subspace

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a + 3b - c - 5d = 0, \ -2a - 6b + 3c + 14d = 0 \right\}$$

and find a spanning set for Z.

We need to construct a limited number of matrices in Z so that every matrix in Z can be expressed as a linear combination of this limited number of matrices. Suppose that

 $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a matrix in Z. Then we can form a column vector with the entries of B and write

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathcal{N}\left(\begin{bmatrix} 1 & 3 & -1 & -5 \\ -2 & -6 & 3 & 14 \end{bmatrix} \right)$$

Row-reducing this matrix and applying Theorem REMES [61] we obtain the equivalent statement,

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \in \mathcal{N}\left(\begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 4 \end{bmatrix} \right)$$

We can then express the subspace Z in the following equal forms,

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a + 3b - c - 5d = 0, \ -2a - 6b + 3c + 14d = 0 \right\}$$
$$= \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a + 3b - d = 0, \ c + 4d = 0 \right\}$$
$$= \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a = -3b + d, \ c = -4d \right\}$$
$$= \left\{ \begin{bmatrix} -3b + d & b \\ -4d & d \end{bmatrix} \middle| b, d \in \mathbb{C} \right\}$$

$$= \left\{ \begin{bmatrix} -3b & b \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} d & 0 \\ -4d & d \end{bmatrix} \middle| b, d \in \mathbb{C} \right\}$$
$$= \left\{ b \begin{bmatrix} -3 & 1 \\ 0 & 0 \end{bmatrix} + d \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \middle| b, d \in \mathbb{C} \right\}$$
$$= \left\langle \left\{ \begin{bmatrix} -3 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \right\} \right\rangle$$

So the set

$$Q = \left\{ \begin{bmatrix} -3 & 1\\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0\\ -4 & 1 \end{bmatrix} \right\}$$

spans Z by Definition TSVS [746].

Example SSC

Spanning set in the crazy vector space

In Example LIC [743] we determined that the set $R = \{(1, 0), (6, 3)\}$ is linearly independent in the crazy vector space C (Example CVS [670]). We now show that R is a spanning set for C.

Given an arbitrary vector $(x, y) \in C$ we desire to show that it can be written as a linear combination of the elements of R. In other words, are there scalars a_1 and a_2 so that

$$(x, y) = a_1(1, 0) + a_2(6, 3)$$

We will act as if this equation is true and try to determine just what a_1 and a_2 would be (as

functions of x and y).

$$\begin{aligned} (x, y) &= a_1(1, 0) + a_2(6, 3) \\ &= (1a_1 + a_1 - 1, 0a_1 + a_1 - 1) + (6a_2 + a_2 - 1, 3a_2 + a_2 - 1) & \text{Scalar mult in } C \\ &= (2a_1 - 1, a_1 - 1) + (7a_2 - 1, 4a_2 - 1) \\ &= (2a_1 - 1 + 7a_2 - 1 + 1, a_1 - 1 + 4a_2 - 1 + 1) & \text{Addition in } C \\ &= (2a_1 + 7a_2 - 1, a_1 + 4a_2 - 1) \end{aligned}$$

Equality in C then yields the two equations,

$$2a_1 + 7a_2 - 1 = x$$
$$a_1 + 4a_2 - 1 = y$$

which becomes the linear system with a matrix representation

$$\begin{bmatrix} 2 & 7 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} x+1 \\ y+1 \end{bmatrix}$$

The coefficient matrix of this system is nonsingular, hence invertible (Theorem NI [542]), and we can employ its inverse to find a solution (Theorem TTMI [509], Theorem SNCM [543]),

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 & 7 \\ 1 & 4 \end{bmatrix}^{-1} \begin{bmatrix} x+1 \\ y+1 \end{bmatrix} = \begin{bmatrix} 4 & -7 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x+1 \\ y+1 \end{bmatrix} = \begin{bmatrix} 4x-7y-3 \\ -x+2y+1 \end{bmatrix}$$

We could chase through the above implications backwards and take the existence of these solutions as sufficient evidence for R being a spanning set for C. Instead, let us view the above as simply scratchwork and now get serious with a simple direct proof that R is a spanning set. Ready? Suppose (x, y) is any vector from C, then compute the following linear combination using the definitions of the operations in C,

$$\begin{array}{l} (4x-7y-3)(1,\,0)+(-x+2y+1)(6,\,3)\\ &=(1(4x-7y-3)+(4x-7y-3)-1,\,0(4x-7y-3)+(4x-7y-3)-1)+\\ &\quad (6(-x+2y+1)+(-x+2y+1)-1,\,3(-x+2y+1)+(-x+2y+1)-1)\\ &=(8x-14y-7,\,4x-7y-4)+(-7x+14y+6,\,-4x+8y+3)\\ &=((8x-14y-7)+(-7x+14y+6)+1,\,(4x-7y-4)+(-4x+8y+3)+1\\ &=(x,\,y) \end{array}$$

This final sequence of computations in C is sufficient to demonstrate that any element of C can be written (or expressed) as a linear combination of the two vectors in R, so $C \subseteq \langle R \rangle$. Since the reverse inclusion $\langle R \rangle \subseteq C$ is trivially true, $C = \langle R \rangle$ and we say R spans C (Definition TSVS [746]). Notice that this demonstration is no more or less valid if we hide from the reader our scratchwork that suggested $a_1 = 4x - 7y - 3$ and $a_2 = -x + 2y + 1$.

Subsection VR Vector Representation

In Chapter R [1260] we will take up the matter of representations fully, where Theorem VRRB [755] will be critical for Definition VR [1261]. We will now motivate and prove a critical theorem that tells us how to "represent" a vector. This theorem could wait, but working with it now will provide some extra insight into the nature of linearly independent spanning sets. First an example, then the theorem.

Example AVR A vector representation Consider the set

$$S = \left\{ \begin{bmatrix} -7\\5\\1 \end{bmatrix}, \begin{bmatrix} -6\\5\\0 \end{bmatrix}, \begin{bmatrix} -12\\7\\4 \end{bmatrix} \right\}$$

from the vector space \mathbb{C}^3 . Let A be the matrix whose columns are the set S, and verify that A is nonsingular. By Theorem NMLIC [330] the elements of S form a linearly independent set. Suppose that $\mathbf{b} \in \mathbb{C}^3$. Then $\mathcal{LS}(A, \mathbf{b})$ has a (unique) solution (Theorem NMUS [180]) and hence is consistent. By Theorem SLSLC [229], $\mathbf{b} \in \langle S \rangle$. Since \mathbf{b} is arbitrary, this is enough to show that $\langle S \rangle = \mathbb{C}^3$, and therefore S is a spanning set for \mathbb{C}^3 (Definition TSVS [746]). (This set comes from the columns of the coefficient matrix of Archetype B [1652].)

Now examine the situation for a particular choice of **b**, say $\mathbf{b} = \begin{bmatrix} -33\\ 24\\ 5 \end{bmatrix}$. Because S is a

spanning set for \mathbb{C}^3 , we know we can write **b** as a linear combination of the vectors in S,

$$\begin{bmatrix} -33\\24\\5 \end{bmatrix} = (-3)\begin{bmatrix} -7\\5\\1 \end{bmatrix} + (5)\begin{bmatrix} -6\\5\\0 \end{bmatrix} + (2)\begin{bmatrix} -12\\7\\4 \end{bmatrix}.$$

The nonsingularity of the matrix A tells that the scalars in this linear combination are unique. More precisely, it is the linear independence of S that provides the uniqueness. We will refer to the scalars $a_1 = -3$, $a_2 = 5$, $a_3 = 2$ as a "representation of **b** relative to S." In other words, once we settle on S as a linearly independent set that spans \mathbb{C}^3 , the vector **b** is recoverable just by knowing the scalars $a_1 = -3$, $a_2 = 5$, $a_3 = 2$ (use these scalars in a linear combination of the vectors in S). This is all an illustration of the following important theorem, which we prove in the setting of a general vector space.

Theorem VRRB

Vector Representation Relative to a Basis

Suppose that V is a vector space and $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m}$ is a linearly independent set that spans V. Let **w** be any vector in V. Then there exist *unique* scalars $a_1, a_2, a_3, \ldots, a_m$ such that

$$\mathbf{w} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_m \mathbf{v}_m.$$

 \square

Proof That \mathbf{w} can be written as a linear combination of the vectors in B follows from the spanning property of the set (Definition TSVS [746]). This is good, but not the meat of this theorem. We now know that for any choice of the vector \mathbf{w} there exist *some* scalars that will create \mathbf{w} as a linear combination of the basis vectors. The real question is: Is there *more* than one way to write \mathbf{w} as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m\}$? Are the scalars $a_1, a_2, a_3, \ldots, a_m$ unique? (Technique U [1625])

Assume there are two ways to express **w** as a linear combination of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m\}$. In other words there exist scalars $a_1, a_2, a_3, \ldots, a_m$ and $b_1, b_2, b_3, \ldots, b_m$ so that

$$\mathbf{w} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_m \mathbf{v}_m$$
$$\mathbf{w} = b_1 \mathbf{v}_1 + b_2 \mathbf{v}_2 + b_3 \mathbf{v}_3 + \dots + b_m \mathbf{v}_m.$$

Then notice that

$$\mathbf{0} = \mathbf{w} + (-\mathbf{w})$$

$$= \mathbf{w} + (-\mathbf{w})$$

$$= (-1)\mathbf{w}$$

$$= (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \dots + a_m\mathbf{v}_m) + (-1)(b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + \dots + b_m\mathbf{v}_m)$$

$$= (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \dots + a_m\mathbf{v}_m) + (-b_1\mathbf{v}_1 - b_2\mathbf{v}_2 - b_3\mathbf{v}_3 - \dots - b_m\mathbf{v}_m)$$

$$= (a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + (a_3 - b_3)\mathbf{v}_3 + \dots$$
Property DVA [660]

$$\cdots + (a_m - b_m)\mathbf{v_m}$$
 Property C [660], Property DSA [660]

But this is a relation of linear dependence on a linearly independent set of vectors (Definition RLD [735])! Now we are using the other assumption about B, that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m\}$ is a linearly independent set. So by Definition LI [735] it *must* happen that the scalars are all zero. That is,

$$(a_1 - b_1) = 0 \qquad (a_2 - b_2) = 0 \qquad (a_3 - b_3) = 0 \qquad \dots \qquad (a_m - b_m) = 0 a_1 = b_1 \qquad a_2 = b_2 \qquad a_3 = b_3 \qquad \dots \qquad a_m = b_m$$

And so we find that the scalars are unique.

This is a very typical use of the hypothesis that a set is linearly independent — obtain a relation of linear dependence and then conclude that the scalars *must* all be zero. The result of this theorem tells us that we can write any vector in a vector space as a linear combination of the vectors in a linearly independent spanning set, but only just. There is only enough raw material in the spanning set to write each vector one way as a linear combination. So in this sense, we could call a linearly independent spanning set a "minimal spanning set." These sets are so important that we will give them a simpler name ("basis") and explore their properties further in the next section.

Subsection READ Reading Questions

1. Is the set of matrices below linearly independent or linearly dependent in the vector space M_{22} ? Why or why not?

$$\left\{ \begin{bmatrix} 1 & 3 \\ -2 & 4 \end{bmatrix}, \begin{bmatrix} -2 & 3 \\ 3 & -5 \end{bmatrix}, \begin{bmatrix} 0 & 9 \\ -1 & 3 \end{bmatrix} \right\}$$

- 2. Explain the difference between the following two uses of the term "span":
 (a) S is a subset of the vector space V and the span of S is a subspace of V.
 (b) W is a subspace of the vector space Y and T spans W.
- 3. The set

$$S = \left\{ \begin{bmatrix} 6\\2\\1 \end{bmatrix}, \begin{bmatrix} 4\\-3\\1 \end{bmatrix}, \begin{bmatrix} 5\\8\\2 \end{bmatrix} \right\}$$

is linearly independent and spans \mathbb{C}^3 . Write the vector $\mathbf{x} = \begin{bmatrix} -6\\ 2\\ 2 \end{bmatrix}$ as a linear combina-

tion of the elements of S. How many ways are there to answer this question, and which theorem allows you to say so?

Subsection EXC Exercises

C20 In the vector space of 2×2 matrices, M_{22} , determine if the set S below is linearly independent.

$$S = \left\{ \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 4 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [764]

C21 In the crazy vector space C (Example CVS [670]), is the set $S = \{(0, 2), (2, 8)\}$ linearly independent?

Contributed by Robert Beezer Solution [765]

C22 In the vector space of polynomials P_3 , determine if the set S is linearly independent or linearly dependent.

$$S = \left\{2 + x - 3x^2 - 8x^3, 1 + x + x^2 + 5x^3, 3 - 4x^2 - 7x^3\right\}$$

Contributed by Robert Beezer Solution [766]

C23 Determine if the set $S = \{(3, 1), (7, 3)\}$ is linearly independent in the crazy vector

space C (Example CVS [670]). Contributed by Robert Beezer Solution [767]

C24 In the vector space of real-valued functions $F = \{f | f : \mathbb{R} \to \mathbb{R}\}$, determine if the following set S is linearly independent.

$$S = \left\{ \sin^2 x, \cos^2 x, 2 \right\}$$

Contributed by Chris Black Solution [767]

C25 Let $S = \left\{ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \right\}$

- 1. Determine if S spans $M_{2,2}$.
- 2. Determine if S is linearly independent.

Contributed by Chris Black Solution [767]

C26 Let

$$S = \left\{ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} \right\}$$

1. Determine if S spans $M_{2,2}$.

2. Determine if S is linearly independent.

Contributed by Chris Black Solution [769]

C30 In Example LIM32 [739], find another nontrivial relation of linear dependence on the linearly dependent set of 3×2 matrices, S. Contributed by Robert Beezer

C40 Determine if the set $T = \{x^2 - x + 5, 4x^3 - x^2 + 5x, 3x + 2\}$ spans the vector space of polynomials with degree 4 or less, P_4 . Contributed by Robert Beezer Solution [772]

C41 The set W is a subspace of M_{22} , the vector space of all 2×2 matrices. Prove that S is a spanning set for W.

$$W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| 2a - 3b + 4c - d = 0 \right\} \qquad S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 4 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [772]

C42 Determine if the set $S = \{(3, 1), (7, 3)\}$ spans the crazy vector space C (Example CVS [670]).

Contributed by Robert Beezer Solution [773]

M10 Halfway through Example SSP4 [746], we need to show that the system of equations

$$\mathcal{LS}\left(\begin{bmatrix} 0 & 0 & 0 & 1\\ 0 & 0 & 1 & -8\\ 0 & 1 & -6 & 24\\ 1 & -4 & 12 & -32\\ -2 & 4 & -8 & 16 \end{bmatrix}, \begin{bmatrix} a\\ b\\ c\\ d\\ e \end{bmatrix}\right)$$

is consistent for every choice of the vector of constants satisfying 16a + 8b + 4c + 2d + e = 0.

Express the column space of the coefficient matrix of this system as a null space, using Theorem FS [625]. From this use Theorem CSCS [563] to establish that the system is always consistent. Notice that this approach removes from Example SSP4 [746] the need to row-reduce a symbolic matrix.

Contributed by Robert Beezer Solution [774]

T40 Prove the following variant of Theorem EMMVP [466] that has a weaker hypothesis: Suppose that $C = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_p}$ is a linearly independent spanning set for \mathbb{C}^n . Suppose also that A and B are $m \times n$ matrices such that $A\mathbf{u}_i = B\mathbf{u}_i$ for every $1 \le i \le n$. Then A = B.

Can you weaken the hypothesis even further while still preserving the conclusion? Contributed by Robert Beezer

T50 Suppose that V is a vector space and $\mathbf{u}, \mathbf{v} \in V$ are two vectors in V. Use the definition of linear independence to prove that $S = {\mathbf{u}, \mathbf{v}}$ is a linearly dependent set if and only if one of the two vectors is a scalar multiple of the other. Prove this directly in the context of

an abstract vector space (V), without simply giving an upgraded version of Theorem DLDS [365] for the special case of just two vectors. Contributed by Robert Beezer Solution [775]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [759]

Begin with a relation of linear dependence on the vectors in S and massage it according to the definitions of vector addition and scalar multiplication in M_{22} ,

$$\mathcal{O} = a_1 \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 4 \\ -1 & 2 \end{bmatrix} + a_3 \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2a_1 + 4a_3 & -a_1 + 4a_2 + 2a_3 \\ a_1 - a_2 + a_3 & 3a_1 + 2a_2 + 3a_3 \end{bmatrix}$$

By our definition of matrix equality (Definition ME [427]) we arrive at a homogeneous system of linear equations,

$$2a_1 + 4a_3 = 0$$

$$-a_1 + 4a_2 + 2a_3 = 0$$

$$a_1 - a_2 + a_3 = 0$$

$$3a_1 + 2a_2 + 3a_3 = 0$$

The coefficient matrix of this system row-reduces to the matrix,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

and from this we conclude that the only solution is $a_1 = a_2 = a_3 = 0$. Since the relation of linear dependence (Definition RLD [735]) is trivial, the set S is linearly independent (Definition LI [735]).

C21 Contributed by Robert Beezer Statement [759]

We begin with a relation of linear dependence using unknown scalars a and b. We wish to know if these scalars *must* both be zero. Recall that the zero vector in C is (-1, -1) and that the definitions of vector addition and scalar multiplication are not what we might expect.

$$\begin{aligned}
\mathbf{0} &= (-1, -1) \\
&= a(0, 2) + b(2, 8) \\
&= (0a + a - 1, 2a + a - 1) + (2b + b - 1, 8b + b - 1) \\
&= (a - 1, 3a - 1) + (3b - 1, 9b - 1) \\
&= (a - 1 + 3b - 1 + 1, 3a - 1 + 9b - 1 + 1) \\
&= (a + 3b - 1, 3a + 9b - 1)
\end{aligned}$$
Definition RLD [735]

Scalar mult., Example CVS [670]

Vector addition, Example CVS [670]

= (a + 3b - 1, 3a + 9b - 1)

From this we obtain two equalities, which can be converted to a homogeneous system of equations,

$$-1 = a + 3b - 1 a + 3b = 0-1 = 3a + 9b - 1 3a + 9b = 0$$

This homogeneous system has a singular coefficient matrix (Theorem SMZD [935]), and so has more than just the trivial solution (Definition NM [173]). Any nontrivial solution will give us a nontrivial relation of linear dependence on S. So S is linearly dependent (Definition LI [735]).

C22 Contributed by Robert Beezer Statement [759] Begin with a relation of linear dependence (Definition RLD [735]),

$$a_1 \left(2 + x - 3x^2 - 8x^3\right) + a_2 \left(1 + x + x^2 + 5x^3\right) + a_3 \left(3 - 4x^2 - 7x^3\right) = \mathbf{0}$$

Massage according to the definitions of scalar multiplication and vector addition in the definition of P_3 (Example VSP [665]) and use the zero vector for this vector space,

$$(2a_1 + a_2 + 3a_3) + (a_1 + a_2)x + (-3a_1 + a_2 - 4a_3)x^2 + (-8a_1 + 5a_2 - 7a_3)x^3 = 0 + 0x + 0x^2 + 0x^3 +$$

The definition of the equality of polynomials allows us to deduce the following four equations,

$$2a_1 + a_2 + 3a_3 = 0$$

$$a_1 + a_2 = 0$$

-3a_1 + a_2 - 4a_3 = 0
-8a_1 + 5a_2 - 7a_3 = 0

Row-reducing the coefficient matrix of this homogeneous system leads to the unique solution $a_1 = a_2 = a_3 = 0$. So the only relation of linear dependence on S is the trivial one, and this is linear independence for S (Definition LI [735]).

C23 Contributed by Robert Beezer Statement [759] Notice, or discover, that the following gives a nontrivial relation of linear dependence on S in C, so by Definition LI [735], the set S is linearly dependent.

$$2(3, 1) + (-1)(7, 3) = (7, 3) + (-9, -5) = (-1, -1) = \mathbf{0}$$

C24 Contributed by Chris Black Statement [760]

One of the fundamental identities of trigonometry is $\sin^2(x) + \cos^2(x) = 1$. Thus, we have a dependence relation $2(\sin^2 x) + 2(\cos^2 x) + (-1)(2) = 0$, and the set is linearly dependent.

C25 Contributed by Chris Black Statement [760]

1. If S spans $M_{2,2}$, then for every 2×2 matrix $B = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$, there exist constants α, β, γ so that

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \alpha \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + \beta \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} + \gamma \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$

Applying Definition ME [427], this leads to the linear system

 α

$$\alpha + 2\beta = x$$

$$2\alpha + \beta + \gamma = y$$

$$2\alpha - \beta + \gamma = z$$

$$+ 2\beta + 2\gamma = w.$$

We need to row-reduce the augmented matrix of this system by hand due to the symbols x, y, z, and w in the vector of constants.

$$\begin{bmatrix} 1 & 2 & 0 & x \\ 2 & 1 & 1 & y \\ 2 & -1 & 1 & z \\ 1 & 2 & 2 & w \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & x - y + z \\ 0 & 1 & 0 & \frac{1}{2}(y - z) \\ 0 & 0 & 1 & \frac{1}{2}(w - x) \\ 0 & 0 & 0 & \frac{1}{2}(5y - 3x - 3z - w) \end{bmatrix}$$

With the apperance of a leading 1 possible in the last column, by Theorem RCLS [120] there will exist some matrices $B = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$ so that the linear system above has no solution (namely, whenever $5y - 3x - 3z - w \neq 0$), so the set S does not span $M_{2,2}$. (For example, you can verify that there is no solution when $B = \begin{bmatrix} 3 & 3 \\ 3 & 2 \end{bmatrix}$.)

2. To check for linear independence, we need to see if there are nontrivial coefficients α, β, γ

that solve

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \alpha \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + \beta \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} + \gamma \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$

This requires the same work that was done in part (a), with x = y = z = w = 0. In that case, the coefficient matrix row-reduces to have a leading 1 in each of the first three columns and a row of zeros on the bottom, so we know that the only solution to the matrix equation is $\alpha = \beta = \gamma = 0$. So the set S is linearly independent.

- C26 Contributed by Chris Black Statement [760]
 - 1. The matrices in S will span $M_{2,2}$ if for any $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$, there are coefficients a, b, c, d, e so that

$$a\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix} + b\begin{bmatrix}2 & 1\\ -1 & 2\end{bmatrix} + c\begin{bmatrix}0 & 1\\ 1 & 2\end{bmatrix} + d\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix} + e\begin{bmatrix}1 & 4\\ 0 & 3\end{bmatrix} = \begin{bmatrix}x & y\\ z & w\end{bmatrix}$$

Thus, we have

$$\begin{bmatrix} a+2b+d+e & 2a+b+c+4e\\ 2a-b+c+d & a+2b+2c+d+3e \end{bmatrix} = \begin{bmatrix} x & y\\ z & w \end{bmatrix}$$

so we have the matrix equation

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & 4 \\ 2 & -1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

This system will have a solution for *every* vector on the right side if the row-reduced coefficient matrix has a leading one in every row, since then it is never possible to have a leading 1 appear in the final column of a row-reduced augmented matrix.

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & 4 \\ 2 & -1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix}$$

Since there is a leading one in each row of the row-reduced coefficient matrix, there is a solution for every vector $\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$, which means that there is a solution to the original equation for every matrix $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$. Thus, the original five matrices span $M_{2,2}$.

г п

2. The matrices in S are linearly independent if the only solution to

$$a \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + b \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} + d \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + e \begin{bmatrix} 1 & 4 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
is $a = b = c = d = e = 0$.
We have

$$\begin{bmatrix} a+2b+d+e & 2a+b+c+4e \\ 2a-b+c+d & a+2b+2c+d+3e \end{bmatrix} = \begin{bmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & 4 \\ 2 & -1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 3 \end{bmatrix} \begin{vmatrix} a \\ b \\ c \\ d \\ e \end{vmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so we need to find the nullspace of the matrix

is

$$\begin{bmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 1 & 1 & 0 & 4 \\ 2 & -1 & 1 & 1 & 0 \\ 1 & 2 & 2 & 1 & 3 \end{bmatrix}$$

We row-reduced this matrix in part (a), and found that there is a column without a leading 1, which correspons to a free variable in a description of the solution set to the homogeneous system, so the nullspace is nontrivial and there are an infinite number of solutions to

$$a\begin{bmatrix}1&2\\2&1\end{bmatrix}+b\begin{bmatrix}2&1\\-1&2\end{bmatrix}+c\begin{bmatrix}0&1\\1&2\end{bmatrix}+d\begin{bmatrix}1&0\\1&1\end{bmatrix}+e\begin{bmatrix}1&4\\0&3\end{bmatrix}=\begin{bmatrix}0&0\\0&0\end{bmatrix}$$

Thus, this set of matrices is not linearly independent.

C40 Contributed by Robert Beezer Statement [761]

The polynomial x^4 is an element of P_4 . Can we write this element as a linear combination of the elements of T? To wit, are there scalars a_1 , a_2 , a_3 such that

$$x^{4} = a_{1} \left(x^{2} - x + 5 \right) + a_{2} \left(4x^{3} - x^{2} + 5x \right) + a_{3} \left(3x + 2 \right)$$

Massaging the right side of this equation, according to the definitions of Example VSP [665], and then equating coefficients, leads to an inconsistent system of equations (check this!). As such, T is not a spanning set for P_4 .

C41 Contributed by Robert Beezer Statement [761]

We want to show that $W = \langle S \rangle$ (Definition TSVS [746]), which is an equality of sets (Definition SE [1604]).

First, show that $\langle S \rangle \subseteq W$. Begin by checking that each of the three matrices in S is a member of the set W. Then, since W is a vector space, the closure properties (Property AC [660], Property SC [660]) guarantee that every linear combination of elements of S remains in W.

Second, show that $W \subseteq \langle S \rangle$. We want to convince ourselves that an arbitrary element of W is a linear combination of elements of S. Choose

$$\mathbf{x} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in W$$

The values of a, b, c, d are not totally arbitrary, since membership in W requires that 2a - 3b + 4c - d = 0. Now, rewrite as follows,

$$\mathbf{x} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$= \begin{bmatrix} a & b \\ c & 2a - 3b + 4c \end{bmatrix}$$

$$= \begin{bmatrix} a & 0 \\ 0 & 2a \end{bmatrix} + \begin{bmatrix} 0 & b \\ 0 & -3b \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ c & 4c \end{bmatrix}$$

$$= a \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 4 \end{bmatrix}$$

$$= a \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 4 \end{bmatrix}$$
Definition MSM [428]
$$\in \langle S \rangle$$
Definition SS [709]

C42 Contributed by Robert Beezer Statement [761] We will try to show that S spans C. Let (x, y) be an arbitrary element of C and search for scalars a_1 and a_2 such that

$$(x, y) = a_1(3, 1) + a_2(7, 3)$$

$$= (4a_1 - 1, 2a_1 - 1) + (8a_2 - 1, 4a_2 - 1)$$
$$= (4a_1 + 8a_2 - 1, 2a_1 + 4a_2 - 1)$$

Equality in C leads to the system

$$4a_1 + 8a_2 = x + 1$$

$$2a_1 + 4a_2 = y + 1$$

This system has a singular coefficient matrix whose column space is simply $\left\langle \begin{bmatrix} 2\\1 \end{bmatrix} \right\rangle$. So any choice of x and y that causes the column vector $\begin{bmatrix} x+1\\y+1 \end{bmatrix}$ to lie outside the column space will lead to an inconsistent system, and hence create an element (x, y) that is not in the span of S. So S does not span C.

For example, choose x = 0 and y = 5, and then we can see that $\begin{bmatrix} 1 \\ 6 \end{bmatrix} \notin \left\langle \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\rangle$ and we know that (0, 5) cannot be written as a linear combination of the vectors in S. A shorter solution might begin by asserting that (0, 5) is not in $\langle S \rangle$ and then establishing this claim alone.

M10 Contributed by Robert Beezer Statement [762] Theorem FS [625] provides the matrix

J

$$L = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \end{bmatrix}$$

and so if A denotes the coefficient matrix of the system, then $\mathcal{C}(A) = \mathcal{N}(L)$. The single homogeneous equation in $\mathcal{LS}(L, \mathbf{0})$ is equivalent to the condition on the vector of constants (use a, b, c, d, e as variables and then multiply by 16).

T50 Contributed by Robert Beezer Statement [762]

 (\Rightarrow) If S is linearly dependent, then there are scalars α and β , not both zero, such that $\alpha \mathbf{u} + \beta \mathbf{v} = \mathbf{0}$. Suppose that $\alpha \neq 0$, the proof proceeds similarly if $\beta \neq 0$. Now,

Property C
Property M
Property S
Property Z
Property A
Definition
Property Z
Property S

D [661] MICN [1598] SMA [660] Z [660] AI [660] LI [735] Z [660]

SMA [660]

which shows that \mathbf{u} is a scalar multiple of \mathbf{v} .

(\Leftarrow) Suppose now that **u** is a scalar multiple of **v**. More precisely, suppose there is a scalar γ such that $\mathbf{u} = \gamma \mathbf{v}$. Then

$$(-1)\mathbf{u} + \gamma \mathbf{v} = (-1)\mathbf{u} + \mathbf{u}$$

= (-1)\mathbf{u} + (1)\mathbf{u} Property O [661]
= ((-1) + 1)\mathbf{u} Property DSA [660]
= 0\mathbf{u} Property AICN [1597]
= \mathbf{0} Theorem ZSSM [676]

This is a relation of linear of linear dependence on S (Definition RLD [735]), which is nontrivial since one of the scalars is -1. Therefore S is linearly dependent by Definition LI [735].

Be careful using this theorem. It is only applicable to sets of two vectors. In particular, linear dependence in a set of three or more vectors can be more complicated than just one vector being a scalar multiple of another.

Section B Bases

A basis of a vector space is one of the most useful concepts in linear algebra. It often provides a concise, finite description of an infinite vector space.

Subsection B Bases

We now have all the tools in place to define a basis of a vector space.

Definition B

Basis

Suppose V is a vector space. Then a subset $S \subseteq V$ is a **basis** of V if it is linearly independent and spans V. \triangle

So, a basis is a linearly independent spanning set for a vector space. The requirement that the set spans V insures that S has enough raw material to build V, while the linear independence requirement insures that we do not have any more raw material than we need. As we shall see soon in Section D [815], a basis is a minimal spanning set.

You may have noticed that we used the term basis for some of the titles of previous theorems (e.g. Theorem BNS [334], Theorem BCS [569], Theorem BRS [583]) and if you review each of these theorems you will see that their conclusions provide linearly independent spanning sets for sets that we now recognize as subspaces of \mathbb{C}^m . Examples associated with these theorems include Example NSLIL [336], Example CSOCD [571] and Example IAS [585]. As we will see, these three theorems will continue to be powerful tools, even in the setting of more general vector spaces.

Furthermore, the archetypes contain an abundance of bases. For each coefficient matrix of a system of equations, and for each archetype defined simply as a matrix, there is a basis for the null space, *three* bases for the column space, and a basis for the row space. For this reason, our subsequent examples will concentrate on bases for vector spaces other than \mathbb{C}^m . Notice that Definition B [777] does not preclude a vector space from having many bases, and this is the case, as hinted above by the statement that the archetypes contain three bases for the column space of a matrix. More generally, we can grab any basis for a vector space, multiply any one basis vector by a non-zero scalar and create a slightly different set that is still a basis. For "important" vector spaces, it will be convenient to have a collection of "nice" bases. When a vector space has a single particularly nice basis, it is sometimes called the **standard basis** though there is nothing precise enough about this term to allow us to define it formally — it is a question of style. Here are some nice bases for important vector spaces.

Theorem SUVB

Standard Unit Vectors are a Basis

The set of standard unit vectors for \mathbb{C}^m (Definition SUV [409]), $B = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_m\} = \{\mathbf{e}_i | 1 \le i \le m\}$ is a basis for the vector space \mathbb{C}^m .

Proof We must show that the set B is both linearly independent and a spanning set for \mathbb{C}^m . First, the vectors in B are, by Definition SUV [409], the columns of the identity matrix, which we know is nonsingular (since it row-reduces to the identity matrix, Theorem NMRRI [176]). And the columns of a nonsingular matrix are linearly independent by Theorem NMLIC [330].

Suppose we grab an arbitrary vector from \mathbb{C}^m , say

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix}$$

Can we write \mathbf{v} as a linear combination of the vectors in B? Yes, and quite simply.

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \dots + v_m \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + v_3 \mathbf{e}_3 + \dots + v_m \mathbf{e}_m$$

this shows that $\mathbb{C}^m \subseteq \langle B \rangle$, which is sufficient to show that B is a spanning set for \mathbb{C}^m .

Example BP

Bases for P_n

The vector space of polynomials with degree at most n, P_n , has the basis

$$B = \{1, x, x^2, x^3, \dots, x^n\}.$$

Another nice basis for P_n is

$$C = \left\{1, 1+x, 1+x+x^{2}, 1+x+x^{2}+x^{3}, \dots, 1+x+x^{2}+x^{3}+\dots+x^{n}\right\}.$$

Checking that each of B and C is a linearly independent spanning set are good exercises. \boxtimes

Example BM

A basis for the vector space of matrices

In the vector space M_{mn} of matrices (Example VSM [663]) define the matrices $B_{k\ell}$, $1 \le k \le m$, $1 \le \ell \le n$ by

$$[B_{k\ell}]_{ij} = \begin{cases} 1 & \text{if } k = i, \ \ell = j \\ 0 & \text{otherwise} \end{cases}$$

So these matrices have entries that are all zeros, with the exception of a lone entry that is one. The set of all mn of them,

$$B = \{ B_{k\ell} | 1 \le k \le m, \ 1 \le \ell \le n \}$$

forms a basis for M_{mn} . See Exercise B.M20 [803].

The bases described above will often be convenient ones to work with. However a basis doesn't have to obviously look like a basis.

Example BSP4

A basis for a subspace of P_4

In Example SSP4 [746] we showed that

$$S = \left\{ x - 2, \ x^2 - 4x + 4, \ x^3 - 6x^2 + 12x - 8, \ x^4 - 8x^3 + 24x^2 - 32x + 16 \right\}$$

is a spanning set for $W = \{p(x) | p \in P_4, p(2) = 0\}$. We will now show that S is also linearly independent in W. Begin with a relation of linear dependence,

$$0 + 0x + 0x^{2} + 0x^{3} + 0x^{4} = \alpha_{1} (x - 2) + \alpha_{2} (x^{2} - 4x + 4) + \alpha_{3} (x^{3} - 6x^{2} + 12x - 8) + \alpha_{4} (x^{4} - 8x^{3} + 24x^{2} - 32x + 16) = \alpha_{4}x^{4} + (\alpha_{3} - 8\alpha_{4}) x^{3} + (\alpha_{2} - 6\alpha_{3} + 24\alpha_{4}) x^{2} + (\alpha_{1} - 4\alpha_{2} + 12\alpha_{3} - 32\alpha_{4}) x + (-2\alpha_{1} + 4\alpha_{2} - 8\alpha_{3} + 16\alpha_{4})$$

Equating coefficients (vector equality in P_4) gives the homogeneous system of five equations in four variables,

$$\alpha_4 = 0$$
$$\alpha_3 - 8\alpha_4 = 0$$

 \boxtimes

$$\alpha_2 - 6\alpha_3 + 24\alpha_4 = 0$$

$$\alpha_1 - 4\alpha_2 + 12\alpha_3 - 32\alpha_4 = 0$$

$$-2\alpha_1 + 4\alpha_2 - 8\alpha_3 + 16\alpha_4 = 0$$

We form the coefficient matrix, and row-reduce to obtain a matrix in reduced row-echelon form

[1]	0	0	0]
0	1	0	0
0	0	1	0
0	0	0	1
0	0	0	0

With *only* the trivial solution to this homogeneous system, we conclude that only scalars that will form a relation of linear dependence are the trivial ones, and therefore the set S is linearly independent (Definition LI [735]). Finally, S has earned the right to be called a basis for W (Definition B [777]).

Example BSM22 A basis for a subspace of M_{22} In Example SSM22 [749] we discovered that

$$Q = \left\{ \begin{bmatrix} -3 & 1\\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0\\ -4 & 1 \end{bmatrix} \right\}$$

is a spanning set for the subspace

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a + 3b - c - 5d = 0, \ -2a - 6b + 3c + 14d = 0 \right\}$$

of the vector space of all 2×2 matrices, M_{22} . If we can also determine that Q is linearly independent in Z (or in M_{22}), then it will qualify as a basis for Z. Let's begin with a relation of linear dependence.

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \alpha_1 \begin{bmatrix} -3 & 1 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -3\alpha_1 + \alpha_2 & \alpha_1 \\ -4\alpha_2 & \alpha_2 \end{bmatrix}$$

Using our definition of matrix equality (Definition ME [427]) we equate corresponding entries and get a homogeneous system of four equations in two variables,

$$-3\alpha_1 + \alpha_2 = 0$$
$$\alpha_1 = 0$$
$$-4\alpha_2 = 0$$
$$\alpha_2 = 0$$

We could row-reduce the coefficient matrix of this homogeneous system, but it is not necessary. The second and fourth equations tell us that $\alpha_1 = 0$, $\alpha_2 = 0$ is the *only* solution to this homogeneous system. This qualifies the set Q as being linearly independent, since the only relation of linear dependence is trivial (Definition LI [735]). Therefore Q is a basis for Z (Definition B [777]).

Example BC Basis for the crazy vector space

In Example LIC [743] and Example SSC [751] we determined that the set $R = \{(1, 0), (6, 3)\}$ from the crazy vector space, C (Example CVS [670]), is linearly independent and is a spanning set for C. By Definition B [777] we see that R is a basis for C.

We have seen that several of the sets associated with a matrix are subspaces of vector spaces of column vectors. Specifically these are the null space (Theorem NSMS [705]), column space (Theorem CSMS [719]), row space (Theorem RSMS [719]) and left null space (Theorem LNSMS [720]). As subspaces they are vector spaces (Definition S [694]) and it is natural to ask about bases for these vector spaces. Theorem BNS [334], Theorem BCS [569], Theorem BRS [583] each have conclusions that provide linearly independent spanning sets for (respectively) the null space, column space, and row space. Notice that each of these theorems contains the word "basis" in its title, even though we did not know the precise meaning of the word at the time. To find a basis for a left null space we can use the definition of this subspace as a null space (Definition LNS [609]) and apply Theorem BNS [334]. Or Theorem FS [625] tells us that the left null space can be expressed as a row space and we can then use Theorem BRS [583].

Theorem BS [376] is another early result that provides a linearly independent spanning set (i.e. a basis) as its conclusion. If a vector space of column vectors can be expressed as a

span of a set of column vectors, then Theorem BS [376] can be employed in a straightforward manner to quickly yield a basis.

Subsection BSCV Bases for Spans of Column Vectors

We have seen several examples of bases in different vector spaces. In this subsection, and the next (Subsection B.BNM [789]), we will consider building bases for \mathbb{C}^m and its subspaces.

Suppose we have a subspace of \mathbb{C}^m that is expressed as the span of a set of vectors, S, and S is not necessarily linearly independent, or perhaps not very attractive. Theorem REMRS [580] says that row-equivalent matrices have identical row spaces, while Theorem BRS [583] says the nonzero rows of a matrix in reduced row-echelon form are a basis for the row space. These theorems together give us a great computational tool for quickly finding a basis for a subspace that is expressed originally as a span.

Example RSB

Row space basis

When we first defined the span of a set of column vectors, in Example SCAD [290] we looked at the set

$$W = \left\langle \left\{ \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 1\\4\\1 \end{bmatrix}, \begin{bmatrix} 7\\-5\\4 \end{bmatrix}, \begin{bmatrix} -7\\-6\\-5 \end{bmatrix} \right\} \right\rangle$$

with an eye towards realizing W as the span of a smaller set. By building relations of linear dependence (though we did not know them by that name then) we were able to remove two vectors and write W as the span of the other two vectors. These two remaining vectors formed a linearly independent set, even though we did not know that at the time.

Now we know that W is a subspace and must have a basis. Consider the matrix, C, whose rows are the vectors in the spanning set for W,

$$C = \begin{bmatrix} 2 & -3 & 1 \\ 1 & 4 & 1 \\ 7 & -5 & 4 \\ -7 & -6 & -5 \end{bmatrix}$$

Then, by Definition RSM [577], the row space of C will be W, $\mathcal{R}(C) = W$. Theorem BRS [583] tells us that if we row-reduce C, the nonzero rows of the row-equivalent matrix in reduced row-echelon form will be a basis for $\mathcal{R}(C)$, and hence a basis for W. Let's do it — C row-reduces to

$$\begin{bmatrix} 1 & 0 & \frac{7}{11} \\ 0 & 1 & \frac{1}{11} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

If we convert the two nonzero rows to column vectors then we have a basis,

$$B = \left\{ \begin{bmatrix} 1\\0\\\frac{7}{11} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{1}{11} \end{bmatrix} \right\}$$

and

$$W = \left\langle \left\{ \begin{bmatrix} 1\\0\\\frac{7}{11} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{1}{11} \end{bmatrix} \right\} \right\rangle$$

For aesthetic reasons, we might wish to multiply each vector in B by 11, which will not change the spanning or linear independence properties of B as a basis. Then we can also write

$$W = \left\langle \left\{ \begin{bmatrix} 11\\0\\7 \end{bmatrix}, \begin{bmatrix} 0\\11\\1 \end{bmatrix} \right\} \right\rangle$$

Example IAS [585] provides another example of this flavor, though now we can notice that X is a subspace, and that the resulting set of three vectors is a basis. This is such a powerful technique that we should do one more example.

Example RS Reducing a span

In Example RSC5 [366] we began with a set of n = 4 vectors from \mathbb{C}^5 ,

$$R = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4\} = \left\{ \begin{bmatrix} 1\\2\\-1\\3\\2 \end{bmatrix}, \begin{bmatrix} 2\\1\\3\\1\\2 \end{bmatrix}, \begin{bmatrix} 0\\-7\\6\\-11\\-2 \end{bmatrix}, \begin{bmatrix} 4\\1\\2\\1\\6 \end{bmatrix} \right\}$$

 \boxtimes

and defined $V = \langle R \rangle$. Our goal in that problem was to find a relation of linear dependence on the vectors in R, solve the resulting equation for one of the vectors, and re-express V as the span of a set of three vectors.

Here is another way to accomplish something similar. The row space of the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 & 3 & 2 \\ 2 & 1 & 3 & 1 & 2 \\ 0 & -7 & 6 & -11 & -2 \\ 4 & 1 & 2 & 1 & 6 \end{bmatrix}$$

is equal to $\langle R \rangle$. By Theorem BRS [583] we can row-reduce this matrix, ignore any zero rows, and use the non-zero rows as column vectors that are a basis for the row space of A. Row-reducing A creates the matrix

$$\begin{bmatrix} 1 & 0 & 0 & -\frac{1}{17} & \frac{30}{17} \\ 0 & 1 & 0 & \frac{25}{17} & -\frac{2}{17} \\ 0 & 0 & 1 & -\frac{2}{17} & -\frac{8}{17} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

 So

$$\left\{ \begin{bmatrix} 1\\0\\0\\-\frac{1}{17}\\\frac{30}{17} \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\\frac{25}{17}\\-\frac{2}{17} \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\-\frac{2}{17}\\-\frac{2}{17} \end{bmatrix} \right\}$$

is a basis for V. Our theorem tells us this is a basis, there is no need to verify that the subspace spanned by three vectors (rather than four) is the identical subspace, and there is no need to verify that we have reached the limit in reducing the set, since the set of three vectors is guaranteed to be linearly independent.

Subsection BNM Bases and Nonsingular Matrices

A quick source of diverse bases for \mathbb{C}^m is the set of columns of a nonsingular matrix.

Theorem CNMB

Columns of Nonsingular Matrix are a Basis

Suppose that A is a square matrix of size m. Then the columns of A are a basis of \mathbb{C}^m if and only if A is nonsingular.

Proof (\Rightarrow) Suppose that the columns of A are a basis for \mathbb{C}^m . Then Definition B [777] says the set of columns is linearly independent. Theorem NMLIC [330] then says that A is nonsingular.

(\Leftarrow) Suppose that A is nonsingular. Then by Theorem NMLIC [330] this set of columns is linearly independent. Theorem CSNM [574] says that for a nonsingular matrix, $C(A) = \mathbb{C}^m$. This is equivalent to saying that the columns of A are a spanning set for the vector space \mathbb{C}^m . As a linearly independent spanning set, the columns of A qualify as a basis for \mathbb{C}^m (Definition B [777]).

Example CABAK Columns as Basis, Archetype K

Archetype K [1734] is the 5×5 matrix

$$K = \begin{bmatrix} 10 & 18 & 24 & 24 & -12 \\ 12 & -2 & -6 & 0 & -18 \\ -30 & -21 & -23 & -30 & 39 \\ 27 & 30 & 36 & 37 & -30 \\ 18 & 24 & 30 & 30 & -20 \end{bmatrix}$$

which is row-equivalent to the 5×5 identity matrix I_5 . So by Theorem NMRRI [176], K is nonsingular. Then Theorem CNMB [789] says the set

$$\left\{ \begin{bmatrix} 10\\12\\-30\\27\\18 \end{bmatrix}, \begin{bmatrix} 18\\-2\\-21\\30\\24 \end{bmatrix}, \begin{bmatrix} 24\\-6\\-23\\36\\30 \end{bmatrix}, \begin{bmatrix} 24\\0\\-30\\37\\30 \end{bmatrix}, \begin{bmatrix} -12\\-18\\39\\-30\\-20 \end{bmatrix} \right\}$$

is a (novel) basis of \mathbb{C}^5 .

Perhaps we should view the fact that the standard unit vectors are a basis (Theorem SUVB [779]) as just a simple corollary of Theorem CNMB [789]? (See Technique LC [1634].)

With a new equivalence for a nonsingular matrix, we can update our list of equivalences.

Theorem NME5 Nonsingular Matrix Equivalences, Round 5

Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.
- 7. The column space of A is \mathbb{C}^n , $\mathcal{C}(A) = \mathbb{C}^n$.
- 8. The columns of A are a basis for \mathbb{C}^n .

Proof With a new equivalence for a nonsingular matrix in Theorem CNMB [789] we can expand Theorem NME4 [575].

 \square

Subsection OBC Orthonormal Bases and Coordinates

We learned about orthogonal sets of vectors in \mathbb{C}^m back in Section O [395], and we also learned that orthogonal sets are automatically linearly independent (Theorem OSLI [412]). When an orthogonal set also spans a subspace of \mathbb{C}^m , then the set is a basis. And when the set is orthonormal, then the set is an incredibly nice basis. We will back up this claim with a theorem, but first consider how you might manufacture such a set.

Suppose that W is a subspace of \mathbb{C}^m with basis B. Then B spans W and is a linearly independent set of nonzero vectors. We can apply the Gram-Schmidt Procedure (Theorem GSP [413]) and obtain a linearly independent set T such that $\langle T \rangle = \langle B \rangle = W$ and T is orthogonal. In other words, T is a basis for W, and is an orthogonal set. By scaling each vector of T to norm 1, we can convert T into an orthonormal set, without destroying the properties that make it a basis of W. In short, we can convert any basis into an orthonormal basis. Example GSTV [417], followed by Example ONTV [418], illustrates this process.

Unitary matrices (Definition UM [544]) are another good source of orthonormal bases (and vice versa). Suppose that Q is a unitary matrix of size n. Then the n columns of Q form an orthonormal set (Theorem CUMOS [547]) that is therefore linearly independent (Theorem OSLI [412]). Since Q is invertible (Theorem UMI [546]), we know Q is nonsingular (Theorem NI [542]), and then the columns of Q span \mathbb{C}^n (Theorem CSNM [574]). So the columns of a unitary matrix of size n are an orthonormal basis for \mathbb{C}^n . Why all the fuss about orthonormal bases? Theorem VRRB [755] told us that any vector in a vector space could be written, uniquely, as a linear combination of basis vectors. For an orthonormal basis, finding the scalars for this linear combination is extremely easy, and this is the content of the next theorem. Furthermore, with vectors written this way (as linear combinations of the elements of an orthonormal set) certain computations and analysis become much easier. Here's the promised theorem.

Theorem COB

Coordinates and Orthonormal Bases

Suppose that $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_p}$ is an orthonormal basis of the subspace W of \mathbb{C}^m . For any $\mathbf{w} \in W$,

$$\mathbf{w} = \langle \mathbf{v}_1, \, \mathbf{w} \rangle \, \mathbf{v}_1 + \langle \mathbf{v}_2, \, \mathbf{w} \rangle \, \mathbf{v}_2 + \langle \mathbf{v}_3, \, \mathbf{w} \rangle \, \mathbf{v}_3 + \dots + \langle \mathbf{v}_p, \, \mathbf{w} \rangle \, \mathbf{v}_p$$

Proof Because *B* is a basis of *W*, Theorem VRRB [755] tells us that we can write **w** uniquely as a linear combination of the vectors in *B*. So it is not this aspect of the conclusion that makes this theorem interesting. What is interesting is that the particular scalars are so easy to compute. No need to solve big systems of equations — just do an inner product of **w** with \mathbf{v}_i to arrive at the coefficient of \mathbf{v}_i in the linear combination.

So begin the proof by writing \mathbf{w} as a linear combination of the vectors in B, using unknown scalars,

$$\mathbf{w} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_p \mathbf{v}_p$$

and compute,

$$\langle \mathbf{v}_{i}, \mathbf{w} \rangle = \left\langle \mathbf{v}_{i}, \sum_{k=1}^{p} a_{k} \mathbf{v}_{k} \right\rangle$$
 Theorem VRRB [755]

$$= \sum_{k=1}^{p} \left\langle \mathbf{v}_{i}, a_{k} \mathbf{v}_{k} \right\rangle$$
 Theorem IPVA [400]

$$= \sum_{k=1}^{p} a_{k} \left\langle \mathbf{v}_{i}, \mathbf{v}_{k} \right\rangle$$
 Theorem IPSM [401]

$$= a_{i} \left\langle \mathbf{v}_{i}, \mathbf{v}_{i} \right\rangle + \sum_{\substack{k=1 \ k \neq i}}^{p} a_{k} \left\langle \mathbf{v}_{i}, \mathbf{v}_{k} \right\rangle$$
 Property C [660]

$$= a_{i}(1) + \sum_{\substack{k=1 \ k \neq i}}^{p} a_{k}(0)$$
 Definition ONS [418]

$$= a_{i}$$

So the (unique) scalars for the linear combination are indeed the inner products advertised in the conclusion of the theorem's statement.

Example CROB4 Coordinatization relative to an orthonormal basis, \mathbb{C}^4

The set

Now,

$$\{\mathbf{x}_{1}, \mathbf{x}_{2}, \mathbf{x}_{3}, \mathbf{x}_{4}\} = \left\{ \begin{bmatrix} 1+i\\1\\1-i\\i \end{bmatrix}, \begin{bmatrix} 1+5i\\6+5i\\-7-i\\1-6i \end{bmatrix}, \begin{bmatrix} -7+34i\\-8-23i\\-10+22i\\30+13i \end{bmatrix}, \begin{bmatrix} -2-4i\\6+i\\4+3i\\6-i \end{bmatrix} \right\}$$

was proposed, and partially verified, as an orthogonal set in Example AOS [410]. Let's scale each vector to norm 1, so as to form an orthonormal set in \mathbb{C}^4 . Then by Theorem OSLI [412] the set will be linearly independent, and by Theorem NME5 [791] the set will be a basis for \mathbb{C}^4 . So, once scaled to norm 1, the adjusted set will be an orthonormal basis of \mathbb{C}^4 . The norms are,

$$\|\mathbf{x}_1\| = \sqrt{6}$$
 $\|\mathbf{x}_2\| = \sqrt{174}$ $\|\mathbf{x}_3\| = \sqrt{3451}$ $\|\mathbf{x}_4\| = \sqrt{119}$

So an orthonormal basis is

$$B = \{\mathbf{v}_{1}, \, \mathbf{v}_{2}, \, \mathbf{v}_{3}, \, \mathbf{v}_{4}\} \\ = \left\{ \frac{1}{\sqrt{6}} \begin{bmatrix} 1+i\\1\\1-i\\i \end{bmatrix}, \, \frac{1}{\sqrt{174}} \begin{bmatrix} 1+5i\\6+5i\\-7-i\\1-6i \end{bmatrix}, \, \frac{1}{\sqrt{3451}} \begin{bmatrix} -7+34i\\-8-23i\\-10+22i\\30+13i \end{bmatrix}, \, \frac{1}{\sqrt{119}} \begin{bmatrix} -2-4i\\6+i\\4+3i\\6-i \end{bmatrix} \right\} \\ \text{to illustrate Theorem COB [793], choose any vector from } \mathbb{C}^{4}, \, \text{say } \mathbf{w} = \begin{bmatrix} 2\\-3\\1\\4 \end{bmatrix}, \text{ and } \begin{bmatrix} 2\\-3\\1\\4 \end{bmatrix}$$

compute

$$\langle \mathbf{w}, \, \mathbf{v}_1 \rangle = \frac{-5i}{\sqrt{6}}, \quad \langle \mathbf{w}, \, \mathbf{v}_2 \rangle = \frac{-19 + 30i}{\sqrt{174}}, \quad \langle \mathbf{w}, \, \mathbf{v}_3 \rangle = \frac{120 - 211i}{\sqrt{3451}}, \quad \langle \mathbf{w}, \, \mathbf{v}_4 \rangle = \frac{6 + 12i}{\sqrt{119}}$$

Then Theorem COB [793] guarantees that

$$\begin{bmatrix} 2\\ -3\\ 1\\ 4 \end{bmatrix} = \frac{-5i}{\sqrt{6}} \left(\frac{1}{\sqrt{6}} \begin{bmatrix} 1+i\\ 1\\ 1-i\\ i \end{bmatrix} \right) + \frac{-19+30i}{\sqrt{174}} \left(\frac{1}{\sqrt{174}} \begin{bmatrix} 1+5i\\ 6+5i\\ -7-i\\ 1-6i \end{bmatrix} \right) + \frac{120-211i}{\sqrt{3451}} \left(\frac{1}{\sqrt{3451}} \begin{bmatrix} -7+34i\\ -8-23i\\ -10+22i\\ 30+13i \end{bmatrix} \right) + \frac{6+12i}{\sqrt{119}} \left(\frac{1}{\sqrt{119}} \begin{bmatrix} -2-4i\\ 6+i\\ 4+3i\\ 6-i \end{bmatrix} \right)$$

as you might want to check (if you have unlimited patience).

A slightly less intimidating example follows, in three dimensions and with just real numbers.

Example CROB3 Coordinatization relative to an orthonormal basis, \mathbb{C}^3 The set

$$\{\mathbf{x}_1, \, \mathbf{x}_2, \, \mathbf{x}_3\} = \left\{ \begin{bmatrix} 1\\2\\1 \end{bmatrix}, \, \begin{bmatrix} -1\\0\\1 \end{bmatrix}, \, \begin{bmatrix} 2\\1\\1 \end{bmatrix} \right\}$$

 \boxtimes

is a linearly independent set, which the Gram-Schmidt Process (Theorem GSP [413]) converts to an orthogonal set, and which can then be converted to the orthonormal set,

$$B = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3\} = \left\{ \frac{1}{\sqrt{6}} \begin{bmatrix} 1\\2\\1 \end{bmatrix}, \, \frac{1}{\sqrt{2}} \begin{bmatrix} -1\\0\\1 \end{bmatrix}, \, \frac{1}{\sqrt{3}} \begin{bmatrix} 1\\-1\\1 \end{bmatrix} \right\}$$

which is therefore an orthonormal basis of \mathbb{C}^3 . With three vectors in \mathbb{C}^3 , all with real number entries, the inner product (Definition IP [398]) reduces to the usual "dot product" (or scalar product) and the orthogonal pairs of vectors can be interpreted as perpendicular pairs of directions. So the vectors in *B* serve as replacements for our usual 3-D axes, or the usual 3-D unit vectors \vec{i}, \vec{j} and \vec{k} . We would like to decompose arbitrary vectors into "components" in the directions of each of these basis vectors. It is Theorem COB [793] that tells us how to do this.

Suppose that we choose $\mathbf{w} = \begin{bmatrix} 2\\ -1\\ 5 \end{bmatrix}$. Compute

$$\langle \mathbf{w}, \mathbf{v}_1 \rangle = \frac{5}{\sqrt{6}}$$
 $\langle \mathbf{w}, \mathbf{v}_2 \rangle = \frac{3}{\sqrt{2}}$ $\langle \mathbf{w}, \mathbf{v}_3 \rangle = \frac{8}{\sqrt{3}}$

then Theorem COB [793] guarantees that

$$\begin{bmatrix} 2\\-1\\5 \end{bmatrix} = \frac{5}{\sqrt{6}} \left(\frac{1}{\sqrt{6}} \begin{bmatrix} 1\\2\\1 \end{bmatrix} \right) + \frac{3}{\sqrt{2}} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} -1\\0\\1 \end{bmatrix} \right) + \frac{8}{\sqrt{3}} \left(\frac{1}{\sqrt{3}} \begin{bmatrix} 1\\-1\\1 \end{bmatrix} \right)$$

which you should be able to check easily, even if you do not have much patience.

Not only do the columns of a unitary matrix form an orthonormal basis, but there is a deeper connection between orthonormal bases and unitary matrices. Informally, the next theorem says that if we transform each vector of an orthonormal basis by multiplying it by a unitary matrix, then the resulting set will be another orthonormal basis. And more remarkably, any matrix with this property must be unitary! As an equivalence (Technique E [1619]) we could take this as our defining property of a unitary matrix, though it might not have the same utility as Definition UM [544].

Theorem UMCOB

Unitary Matrices Convert Orthonormal Bases

Let A be an $n \times n$ matrix and $B = {\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n}$ be an orthonormal basis of \mathbb{C}^n . Define

$$C = \{A\mathbf{x}_1, A\mathbf{x}_2, A\mathbf{x}_3, \dots, A\mathbf{x}_n\}$$

Then A is a unitary matrix if and only if C is an orthonormal basis of \mathbb{C}^n .

Proof (\Rightarrow) Assume A is a unitary matrix and establish several facts about C. First we check that C is an orthonormal set (Definition ONS [418]). By Theorem UMPIP [549], for $i \neq j$,

$$\langle A\mathbf{x}_i, A\mathbf{x}_j \rangle = \langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$$

Similarly, Theorem UMPIP [549] also gives, for $1 \le i \le n$,

$$\|A\mathbf{x}_i\| = \|\mathbf{x}_i\| = 1$$

 \boxtimes

As C is an orthogonal set (Definition OSV [409]), Theorem OSLI [412] yields the linear independence of C. Having established that the column vectors on C form a linearly independent set, a matrix whose columns are the vectors of C is nonsingular (Theorem NMLIC [330]), and hence these vectors form a basis of \mathbb{C}^n by Theorem CNMB [789].

(\Leftarrow) Now assume that C is an orthonormal set. Let **y** be an arbitrary vector from \mathbb{C}^n . Since B spans \mathbb{C}^n , there are scalars, $a_1, a_2, a_3, \ldots, a_n$, such that

$$\mathbf{y} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 + \dots + a_n \mathbf{x}_n$$

Now

$$A^*A\mathbf{y} = \sum_{i=1}^n \langle \mathbf{x}_i, A^*A\mathbf{y} \rangle \mathbf{x}_i \qquad \text{Theorem COB [793]}$$
$$= \sum_{i=1}^n \left\langle \mathbf{x}_i, A^*A\sum_{j=1}^n a_j \mathbf{x}_j \right\rangle \mathbf{x}_i \qquad \text{Definition TSVS [746]}$$
$$= \sum_{i=1}^n \left\langle \mathbf{x}_i, \sum_{j=1}^n A^*Aa_j \mathbf{x}_j \right\rangle \mathbf{x}_i \qquad \text{Theorem MMDAA [476]}$$
$$= \sum_{i=1}^n \left\langle \mathbf{x}_i, \sum_{j=1}^n a_j A^*A\mathbf{x}_j \right\rangle \mathbf{x}_i \qquad \text{Theorem MMSMM [477]}$$
$$= \sum_{i=1}^n \sum_{j=1}^n \langle \mathbf{x}_i, a_j A^*A\mathbf{x}_j \rangle \mathbf{x}_i \qquad \text{Theorem IPVA [400]}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{j} \langle \mathbf{x}_{i}, A^{*}A\mathbf{x}_{j} \rangle \mathbf{x}_{i}$$
 Theorem IPSM [401]

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{j} \langle A\mathbf{x}_{i}, A\mathbf{x}_{j} \rangle \mathbf{x}_{i}$$
 Theorem AIP [485]

$$= \sum_{i=1}^{n} \sum_{\substack{j=1\\j\neq i}}^{n} a_{j} \langle A\mathbf{x}_{i}, A\mathbf{x}_{j} \rangle \mathbf{x}_{i} + \sum_{\ell=1}^{n} a_{\ell} \langle A\mathbf{x}_{\ell}, A\mathbf{x}_{\ell} \rangle \mathbf{x}_{\ell}$$
 Property C [660]

$$= \sum_{i=1}^{n} \sum_{\substack{j=1\\j\neq i}}^{n} a_{j}(0)\mathbf{x}_{i} + \sum_{\ell=1}^{n} a_{\ell}(1)\mathbf{x}_{\ell}$$
 Definition ONS [418]

$$= \sum_{i=1}^{n} \sum_{\substack{j=1\\j\neq i}}^{n} \mathbf{0} + \sum_{\ell=1}^{n} a_{\ell}\mathbf{x}_{\ell}$$
 Theorem ZSSM [676]

$$= \sum_{\ell=1}^{n} a_{\ell}\mathbf{x}_{\ell}$$
 Property Z [660]

$$= \mathbf{y}$$

$$= I_{n}\mathbf{y}$$
 Theorem MMIM [475]

Since the choice of **y** was arbitrary, Theorem EMMVP [466] tells us that $A^*A = I_n$, so A is unitary (Definition UM [544]).

Subsection READ Reading Questions

1. The matrix below is nonsingular. What can you now say about its columns?

$$A = \begin{bmatrix} -3 & 0 & 1\\ 1 & 2 & 1\\ 5 & 1 & 6 \end{bmatrix}$$

- 2. Write the vector $\mathbf{w} = \begin{bmatrix} 6\\ 6\\ 15 \end{bmatrix}$ as a linear combination of the columns of the matrix A above. How many ways are there to answer this question?
- 3. Why is an orthonormal basis desirable?

Subsection EXC Exercises

C10 Find a basis for $\langle S \rangle$, where

$$S = \left\{ \begin{bmatrix} 1\\3\\2\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\1\\1 \end{bmatrix}, \begin{bmatrix} 1\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\2\\1 \end{bmatrix}, \begin{bmatrix} 3\\4\\1\\3 \end{bmatrix} \right\}.$$

Contributed by Chris Black Solution [805]

C11 Find a basis for the subspace W of \mathbb{C}^4 ,

$$W = \left\{ \begin{bmatrix} a+b-2c\\a+b-2c+d\\-2a+2b+4c-d\\b+d \end{bmatrix} \middle| a,b,c,d \in \mathbb{C} \right\}$$

Contributed by Chris Black Solution [806]

C12 Find a basis for the vector space T of lower triangular 3×3 matrices; that is, matrices

of the form $\begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}$ where an asterisk represents any complex number. Contributed by Chris Black Solution [807]

C13 Find a basis for the subspace Q of P_2 , defined by $Q = \{p(x) = a + bx + cx^2 | p(0) = 0\}.$

Contributed by Chris Black Solution [808]

C14 Find a basis for the subspace R of P_2 defined by $R = \{p(x) = a + bx + cx^2 | p'(0) = 0\}$, where p' denotes the derivative. Contributed by Chris Black Solution [808]

C40 From Example RSB [785], form an arbitrary (and nontrivial) linear combination of the four vectors in the original spanning set for W. So the result of this computation is of course an element of W. As such, this vector should be a linear combination of the basis vectors in B. Find the (unique) scalars that provide this linear combination. Repeat with another linear combination of the original four vectors. Contributed by Robert Beezer Solution [810]

C80 Prove that $\{(1, 2), (2, 3)\}$ is a basis for the crazy vector space C (Example CVS [670]).

Contributed by Robert Beezer

M20 In Example BM [780] provide the verifications (linear independence and spanning)

to show that B is a basis of M_{mn} . Contributed by Robert Beezer Solution [808]

T50 Theorem UMCOB [798] says that unitary matrices are characterized as those matrices that "carry" orthonormal bases to orthonormal bases. This problem asks you to prove a similar result: nonsingular matrices are characterized as those matrices that "carry" bases to bases.

More precisely, suppose that A is a square matrix of size n and $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$ is a basis of \mathbb{C}^n . Prove that A is nonsingular if and only if $C = \{A\mathbf{x}_1, A\mathbf{x}_2, A\mathbf{x}_3, \dots, A\mathbf{x}_n\}$ is a basis of \mathbb{C}^n . (See also Exercise PD.T33 [877], Exercise MR.T20 [1336].) Contributed by Robert Beezer Solution [811]

T51 Use the result of Exercise B.T50 [804] to build a very concise proof of Theorem CNMB [789]. (Hint: make a judicious choice for the basis B.) Contributed by Robert Beezer Solution [814]

Subsection SOL Solutions

C10 Contributed by Chris Black Statement [802]

Theorem BS [376] says that if we take these 5 vectors, put them into a matrix, and rowreduce to discover the pivot columns, then the corresponding vectors in S will be linearly independent and span S, and thus will form a basis of S.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 3 \\ 3 & 2 & 1 & 2 & 4 \\ 2 & 1 & 0 & 2 & 1 \\ 1 & 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 0 & -2 \\ 0 & 1 & 2 & 0 & 5 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, the independent vectors that span S are the first, second and fourth of the set, so a basis of S is

$$B = \left\{ \begin{bmatrix} 1\\3\\2\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\1\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\2\\1 \end{bmatrix} \right\}$$

C11 Contributed by Chris Black Statement [802]

We can rewrite an arbitrary vector of W as

$$\begin{bmatrix} a+b-2c\\ a+b-2c+d\\ -2a+2b+4c-d\\ b+d \end{bmatrix} = \begin{bmatrix} a\\ a\\ -2a\\ 0 \end{bmatrix} + \begin{bmatrix} b\\ b\\ 2b\\ b \end{bmatrix} + \begin{bmatrix} -2c\\ -2c\\ 4c\\ 0 \end{bmatrix} + \begin{bmatrix} 0\\ d\\ -d\\ d \end{bmatrix}$$
$$= a\begin{bmatrix} 1\\ 1\\ -2\\ 0 \end{bmatrix} + b\begin{bmatrix} 1\\ 1\\ 2\\ 1 \end{bmatrix} + c\begin{bmatrix} -2\\ 4c\\ 0 \end{bmatrix} + d\begin{bmatrix} 0\\ 1\\ -1\\ 1 \end{bmatrix}$$

Thus, we can write W as

$$W = \left\langle \left\{ \begin{bmatrix} 1\\1\\-2\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\2\\1 \end{bmatrix}, \begin{bmatrix} -2\\-2\\4\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\-1\\1 \end{bmatrix} \right\} \right\rangle$$

These four vectors span W, but we also need to determine if they are linearly independent (turns out they are not). With an application of Theorem BS [376] we can see that the arrive at a basis employing three of these vectors,

$$\begin{bmatrix} 1 & 1 & -2 & 0 \\ 1 & 1 & -2 & 1 \\ -2 & 2 & 4 & -1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, we have the following basis of W,

$$B = \left\{ \begin{bmatrix} 1\\1\\-2\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\2\\1 \end{bmatrix}, \begin{bmatrix} 0\\1\\-1\\1 \end{bmatrix} \right\}$$

C12 Contributed by Chris Black Statement [802] Let A be an arbitrary element of the specified vector space T. Then there exist a, b, c, d, e and f so that $A = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}$. Then $A = a \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} + e \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} + f \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ Consider the set $B = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 &$

The six vectors in B span the vector space T, and we can check rather simply that they are also linearly independent. Thus, B is a basis of T.

C13 Contributed by Chris Black Statement [803]

If p(0) = 0, then $a+b(0)+c(0^2) = 0$, so a = 0. Thus, we can write $Q = \{p(x) = bx + cx^2 | b, c \in \mathbb{C}\}$. A linearly independent set that spans Q is $B = \{x, x^2\}$, and this set forms a basis of Q.

C14 Contributed by Chris Black Statement [803] The derivative of $p(x) = a+bx+cx^2$ is p'(x) = b+2cx. Thus, if $p \in R$, then p'(0) = b+2c(0) = 0, so we must have b = 0. We see that we can rewrite R as $R = \{p(x) = a + cx^2 | a, c \in \mathbb{C}\}$. A linearly independent set that spans R is $B = \{1, x^2\}$, and B is a basis of R.

M20 Contributed by Robert Beezer Statement [803] We need to establish the linear independence and spanning properties of the set

$$B = \{ B_{k\ell} | 1 \le k \le m, \ 1 \le \ell \le n \}$$

relative to the vector space M_{mn} .

This proof is more transparent if you write out individual matrices in the basis with lots of zeros and dots and a lone one. But we don't have room for that here, so we will use summation notation. Think carefully about each step, especially when the double summations seem to "disappear." Begin with a relation of linear dependence, using double subscripts on the scalars to align with the basis elements.

$$\mathcal{O} = \sum_{k=1}^{m} \sum_{\ell=1}^{n} \alpha_{k\ell} B_{k\ell}$$

Now consider the entry in row i and column j for these equal matrices,

 $0 = [\mathcal{O}]_{ij} \qquad \qquad \text{Definition ZM [433]}$

$$= \left[\sum_{k=1}^{m} \sum_{\ell=1}^{n} \alpha_{k\ell} B_{k\ell}\right]_{ij}$$
 Definition ME [427]
$$= \sum_{k=1}^{m} \sum_{\ell=1}^{n} [\alpha_{k\ell} B_{k\ell}]_{ij}$$
 Definition MA [427]
$$= \sum_{k=1}^{m} \sum_{\ell=1}^{n} \alpha_{k\ell} [B_{k\ell}]_{ij}$$
 Definition MSM [428]
$$= \alpha_{ij} [B_{ij}]_{ij}$$
 [$B_{k\ell}]_{ij} = 0$ when $(k, \ell) \neq (i, j)$
$$= \alpha_{ij}$$

Since i and j were arbitrary, we find that each scalar is zero and so B is linearly independent (Definition LI [735]).

To establish the spanning property of B we need only show that an arbitrary matrix A can be written as a linear combination of the elements of B. So suppose that A is an arbitrary $m \times n$ matrix and consider the matrix C defined as a linear combination of the elements of B by

$$C = \sum_{k=1}^{m} \sum_{\ell=1}^{n} \left[A\right]_{k\ell} B_{k\ell}$$

Then,

$$[C]_{ij} = \left[\sum_{k=1}^{m} \sum_{\ell=1}^{n} [A]_{k\ell} B_{k\ell}\right]_{ij} \qquad \text{Def}$$
$$= \sum_{k=1}^{m} \sum_{\ell=1}^{n} [[A]_{k\ell} B_{k\ell}]_{ij} \qquad \text{Def}$$
$$= \sum_{k=1}^{m} \sum_{\ell=1}^{n} [A]_{k\ell} [B_{k\ell}]_{ij} \qquad \text{Def}$$
$$= [A]_{ij} [B_{ij}]_{ij} \qquad [B_{k\ell}]_{ij} \qquad [B_{k\ell}]_{ij}$$
$$= [A]_{ij} (1) \qquad [B_{ij}]_{ij} \qquad [B_{k\ell}]_{ij}$$

Definition ME [427]

Definition MA [427]

Definition MSM [428]

$$[B_{k\ell}]_{ij} = 0 \text{ when } (k, \ell) \neq (i, j)$$
$$[B_{ij}]_{ij} = 1$$

So by Definition ME [427], A = C, and therefore $A \in \langle B \rangle$. By Definition B [777], the set B is a basis of the vector space M_{mn} .

C40 Contributed by Robert Beezer Statement [803] An arbitrary linear combination is

$$\mathbf{y} = 3 \begin{bmatrix} 2\\ -3\\ 1 \end{bmatrix} + (-2) \begin{bmatrix} 1\\ 4\\ 1 \end{bmatrix} + 1 \begin{bmatrix} 7\\ -5\\ 4 \end{bmatrix} + (-2) \begin{bmatrix} -7\\ -6\\ -5 \end{bmatrix} = \begin{bmatrix} 25\\ -10\\ 15 \end{bmatrix}$$

(You probably used a different collection of scalars.) We want to write ${\bf y}$ as a linear combination of

$$B = \left\{ \begin{bmatrix} 1\\0\\\frac{7}{11} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{1}{11} \end{bmatrix} \right\}$$

We could set this up as vector equation with variables as scalars in a linear combination of the vectors in B, but since the first two slots of B have such a nice pattern of zeros and ones, we can determine the necessary scalars easily and then double-check our answer with a computation in the third slot,

$$25\begin{bmatrix}1\\0\\\frac{7}{11}\end{bmatrix} + (-10)\begin{bmatrix}0\\1\\\frac{1}{11}\end{bmatrix} = \begin{bmatrix}25\\-10\\(25)\frac{7}{11} + (-10)\frac{1}{11}\end{bmatrix} = \begin{bmatrix}25\\-10\\15\end{bmatrix} = \mathbf{y}$$

Notice how the uniqueness of these scalars arises. They are *forced* to be 25 and -10.

T50 Contributed by Robert Beezer Statement [804]

Our first proof relies mostly on definitions of linear independence and spanning, which is a good exercise. The second proof is shorter and turns on a technical result from our work with matrix inverses, Theorem NPNT [538].

 (\Rightarrow) Assume that A is nonsingular and prove that C is a basis of \mathbb{C}^n . First show that C is linearly independent. Work on a relation of linear dependence on C,

$$\mathbf{0} = a_1 A \mathbf{x}_1 + a_2 A \mathbf{x}_2 + a_3 A \mathbf{x}_3 + \dots + a_n A \mathbf{x}_n \qquad \text{Definition RLD [735]}$$

$$= Aa_1\mathbf{x}_1 + Aa_2\mathbf{x}_2 + Aa_3\mathbf{x}_3 + \dots + Aa_n\mathbf{x}_n$$
 Theorem MMSMM [477]
$$= A(a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \dots + a_n\mathbf{x}_n)$$
 Theorem MMDAA [476]

Since A is nonsingular, Definition NM [173] and Theorem SLEMM [462] allows us to conclude that

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \dots + a_n\mathbf{x}_n = \mathbf{0}$$

But this is a relation of linear dependence of the linearly independent set B, so the scalars are trivial, $a_1 = a_2 = a_3 = \cdots = a_n = 0$. By Definition LI [735], the set C is linearly independent.

Now prove that C spans \mathbb{C}^n . Given an arbitrary vector $\mathbf{y} \in \mathbb{C}^n$, can it be expressed as a linear combination of the vectors in C? Since A is a nonsingular matrix we can define the vector \mathbf{w} to be the unique solution of the system $\mathcal{LS}(A, \mathbf{y})$ (Theorem NMUS [180]). Since $\mathbf{w} \in \mathbb{C}^n$ we can write \mathbf{w} as a linear combination of the vectors in the basis B. So there are scalars, $b_1, b_2, b_3, \ldots, b_n$ such that

$$\mathbf{w} = b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + b_3 \mathbf{x}_3 + \dots + b_n \mathbf{x}_n$$

Then,

$$\mathbf{y} = A\mathbf{w}$$

= $A (b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + b_3\mathbf{x}_3 + \dots + b_n\mathbf{x}_n)$
= $Ab_1\mathbf{x}_1 + Ab_2\mathbf{x}_2 + Ab_3\mathbf{x}_3 + \dots + Ab_n\mathbf{x}_n$

Theorem SLEMM [462] Definition TSVS [746] Theorem MMDAA [476]

$$= b_1 A \mathbf{x}_1 + b_2 A \mathbf{x}_2 + b_3 A \mathbf{x}_3 + \dots + b_n A \mathbf{x}_n \qquad \text{Theorem MMSMM [477]}$$

So we can write an arbitrary vector of \mathbb{C}^n as a linear combination of the elements of C. In other words, C spans \mathbb{C}^n (Definition TSVS [746]). By Definition B [777], the set C is a basis for \mathbb{C}^n .

(\Leftarrow) Assume that *C* is a basis and prove that *A* is nonsingular. Let **x** be a solution to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$. Since *B* is a basis of \mathbb{C}^n there are scalars, $a_1, a_2, a_3, \ldots, a_n$, such that

$$\mathbf{x} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 + \dots + a_n \mathbf{x}_n$$

Then

$$\mathbf{0} = A\mathbf{x}$$

$$= A (a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \dots + a_n\mathbf{x}_n)$$

$$= Aa_1\mathbf{x}_1 + Aa_2\mathbf{x}_2 + Aa_3\mathbf{x}_3 + \dots + Aa_n\mathbf{x}_n$$

$$= a_1A\mathbf{x}_1 + a_2A\mathbf{x}_2 + a_3A\mathbf{x}_3 + \dots + a_nA\mathbf{x}_n$$
Theorem MMDAA [476]
Theorem MMSMM [477]

This is a relation of linear dependence on the linearly independent set C, so the scalars must all be zero, $a_1 = a_2 = a_3 = \cdots = a_n = 0$. Thus,

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \dots + a_n\mathbf{x}_n = 0\mathbf{x}_1 + 0\mathbf{x}_2 + 0\mathbf{x}_3 + \dots + 0\mathbf{x}_n = \mathbf{0}.$$

By Definition NM [173] we see that A is nonsingular.

Now for a second proof. Take the vectors for B and use them as the columns of a matrix, $G = [\mathbf{x}_1 | \mathbf{x}_2 | \mathbf{x}_3 | \dots | \mathbf{x}_n]$. By Theorem CNMB [789], because we have the hypothesis that Bis a basis of \mathbb{C}^n , G is a nonsingular matrix. Notice that the columns of AG are exactly the vectors in the set C, by Definition MM [468].

A nonsingular \longleftrightarrow AG nonsingularTheorem NPNT [538] \iff C basis for \mathbb{C}^n Theorem CNMB [789]

That was easy!

T51 Contributed by Robert Beezer Statement [804] Choose *B* to be the set of standard unit vectors, a particularly nice basis of \mathbb{C}^n (Theorem SUVB [779]). For a vector \mathbf{e}_i (Definition SUV [409]) from this basis, what is $A\mathbf{e}_i$?

Section D Dimension

Almost every vector space we have encountered has been infinite in size (an exception is Example VSS [668]). But some are bigger and richer than others. Dimension, once suitably defined, will be a measure of the size of a vector space, and a useful tool for studying its properties. You probably already have a rough notion of what a mathematical definition of dimension might be — try to forget these imprecise ideas and go with the new ones given here.

Subsection D Dimension

Definition D Dimension

Suppose that V is a vector space and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_t\}$ is a basis of V. Then the **dimension** of V is defined by dim (V) = t. If V has no finite bases, we say V has infinite dimension.

(This definition contains Notation D.)

This is a very simple definition, which belies its power. Grab a basis, any basis, and count up the number of vectors it contains. That's the dimension. However, this simplicity causes a problem. Given a vector space, you and I could each construct different bases — remember that a vector space might have many bases. And what if your basis and my basis had different sizes? Applying Definition D [815] we would arrive at different numbers! With our current knowledge about vector spaces, we would have to say that dimension is not "well-defined." Fortunately, there is a theorem that will correct this problem.

In a strictly logical progression, the next two theorems would *precede* the definition of dimension. Many subsequent theorems will trace their lineage back to the following fundamental result.

Theorem SSLD

Spanning Sets and Linear Dependence

Suppose that $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_t}$ is a finite set of vectors which spans the vector space V. Then any set of t + 1 or more vectors from V is linearly dependent. \Box

Proof We want to prove that any set of t+1 or more vectors from V is linearly dependent. So we will begin with a totally arbitrary set of vectors from V, $R = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_m}$, where m > t. We will now construct a nontrivial relation of linear dependence on R.

Each vector \mathbf{u}_1 , \mathbf{u}_2 , \mathbf{u}_3 , ..., \mathbf{u}_m can be written as a linear combination of \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3 , ..., \mathbf{v}_t since S is a spanning set of V. This means there exist scalars a_{ij} , $1 \le i \le t$, $1 \le j \le m$, so that

$$\mathbf{u}_1 = a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3 + \dots + a_{t1}\mathbf{v}_t$$

 \triangle

$$\mathbf{u}_{2} = a_{12}\mathbf{v}_{1} + a_{22}\mathbf{v}_{2} + a_{32}\mathbf{v}_{3} + \dots + a_{t2}\mathbf{v}_{t}$$
$$\mathbf{u}_{3} = a_{13}\mathbf{v}_{1} + a_{23}\mathbf{v}_{2} + a_{33}\mathbf{v}_{3} + \dots + a_{t3}\mathbf{v}_{t}$$
$$\vdots$$
$$\mathbf{u}_{m} = a_{1m}\mathbf{v}_{1} + a_{2m}\mathbf{v}_{2} + a_{3m}\mathbf{v}_{3} + \dots + a_{tm}\mathbf{v}_{t}$$

Now we form, unmotivated, the homogeneous system of t equations in the m variables, $x_1, x_2, x_3, \ldots, x_m$, where the coefficients are the just-discovered scalars a_{ij} ,

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1m}x_m = 0$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2m}x_m = 0$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \dots + a_{3m}x_m = 0$$

$$\vdots$$

$$a_{t1}x_1 + a_{t2}x_2 + a_{t3}x_3 + \dots + a_{tm}x_m = 0$$

This is a homogeneous system with more variables than equations (our hypothesis is expressed as m > t), so by Theorem HMVEI [151] there are infinitely many solutions. Choose a nontrivial solution and denote it by $x_1 = c_1, x_2 = c_2, x_3 = c_3, \ldots, x_m = c_m$. As a solution to the homogeneous system, we then have

$$a_{11}c_1 + a_{12}c_2 + a_{13}c_3 + \dots + a_{1m}c_m = 0$$

$$a_{21}c_1 + a_{22}c_2 + a_{23}c_3 + \dots + a_{2m}c_m = 0$$

$$a_{31}c_1 + a_{32}c_2 + a_{33}c_3 + \dots + a_{3m}c_m = 0$$

$$\vdots$$

$$a_{t1}c_1 + a_{t2}c_2 + a_{t3}c_3 + \dots + a_{tm}c_m = 0$$

As a collection of nontrivial scalars, $c_1, c_2, c_3, \ldots, c_m$ will provide the nontrivial relation of linear dependence we desire,

$$c_{1}\mathbf{u}_{1} + c_{2}\mathbf{u}_{2} + c_{3}\mathbf{u}_{3} + \dots + c_{m}\mathbf{u}_{m}$$

$$= c_{1} (a_{11}\mathbf{v}_{1} + a_{21}\mathbf{v}_{2} + a_{31}\mathbf{v}_{3} + \dots + a_{t1}\mathbf{v}_{t}) \qquad \text{Definition TSVS [746]}$$

$$+ c_{2} (a_{12}\mathbf{v}_{1} + a_{22}\mathbf{v}_{2} + a_{32}\mathbf{v}_{3} + \dots + a_{t2}\mathbf{v}_{t})$$

$$+ c_{3} (a_{13}\mathbf{v}_{1} + a_{23}\mathbf{v}_{2} + a_{33}\mathbf{v}_{3} + \dots + a_{t3}\mathbf{v}_{t})$$

$$\vdots$$

$$+ c_{m} (a_{1m}\mathbf{v}_{1} + a_{2m}\mathbf{v}_{2} + a_{3m}\mathbf{v}_{3} + \dots + a_{tm}\mathbf{v}_{t})$$

$$= c_{1}a_{11}\mathbf{v}_{1} + c_{1}a_{21}\mathbf{v}_{2} + c_{1}a_{31}\mathbf{v}_{3} + \dots + c_{1}a_{t1}\mathbf{v}_{t} \qquad \text{Property DVA [660]}$$

$$+ c_{2}a_{12}\mathbf{v}_{1} + c_{2}a_{22}\mathbf{v}_{2} + c_{2}a_{32}\mathbf{v}_{3} + \dots + c_{2}a_{t2}\mathbf{v}_{t}$$

$$+ c_{3}a_{13}\mathbf{v}_{1} + c_{3}a_{23}\mathbf{v}_{2} + c_{3}a_{33}\mathbf{v}_{3} + \dots + c_{3}a_{t3}\mathbf{v}_{t}$$

$$\vdots$$

$$+ c_{m}a_{1m}\mathbf{v}_{1} + c_{m}a_{2m}\mathbf{v}_{2} + c_{m}a_{3m}\mathbf{v}_{3} + \dots + c_{m}a_{tm}\mathbf{v}_{t}$$

$$= (c_{1}a_{11} + c_{2}a_{12} + c_{3}a_{13} + \dots + c_{m}a_{1m})\mathbf{v}_{1} \qquad \text{Property DSA [660]}$$

$$+ (c_{1}a_{21} + c_{2}a_{22} + c_{3}a_{23} + \dots + c_{m}a_{2m}) \mathbf{v}_{2} + (c_{1}a_{31} + c_{2}a_{32} + c_{3}a_{33} + \dots + c_{m}a_{3m}) \mathbf{v}_{3} \vdots + (c_{1}a_{t1} + c_{2}a_{t2} + c_{3}a_{t3} + \dots + c_{m}a_{tm}) \mathbf{v}_{t} = (a_{11}c_{1} + a_{12}c_{2} + a_{13}c_{3} + \dots + a_{1m}c_{m}) \mathbf{v}_{1} + (a_{21}c_{1} + a_{22}c_{2} + a_{23}c_{3} + \dots + a_{2m}c_{m}) \mathbf{v}_{2} + (a_{31}c_{1} + a_{32}c_{2} + a_{33}c_{3} + \dots + a_{3m}c_{m}) \mathbf{v}_{3} \vdots + (a_{t1}c_{1} + a_{t2}c_{2} + a_{t3}c_{3} + \dots + a_{tm}c_{m}) \mathbf{v}_{t} = 0 \mathbf{v}_{1} + 0 \mathbf{v}_{2} + 0 \mathbf{v}_{3} + \dots + 0 \mathbf{v}_{t} = 0 + 0 + 0 + \dots + 0$$

= 0
Property Z [660]

That does it. R has been undeniably shown to be a linearly dependent set.

The proof just given has some monstrous expressions in it, mostly owing to the double subscripts present. Now is a great opportunity to show the value of a more compact notation. We will rewrite the key steps of the previous proof using summation notation, resulting in a more economical presentation, and even greater insight into the key aspects of the proof. So here is an alternate proof — study it carefully.

Proof (Alternate Proof of Theorem SSLD) We want to prove that any set of t + 1

or more vectors from V is linearly dependent. So we will begin with a totally arbitrary set of vectors from $V, R = \{\mathbf{u}_j | 1 \le j \le m\}$, where m > t. We will now construct a nontrivial relation of linear dependence on R.

Each vector \mathbf{u}_j , $1 \leq j \leq m$ can be written as a linear combination of \mathbf{v}_i , $1 \leq i \leq t$ since S is a spanning set of V. This means there are scalars a_{ij} , $1 \leq i \leq t$, $1 \leq j \leq m$, so that

$$\mathbf{u}_j = \sum_{i=1}^t a_{ij} \mathbf{v}_i \qquad \qquad 1 \le j \le m$$

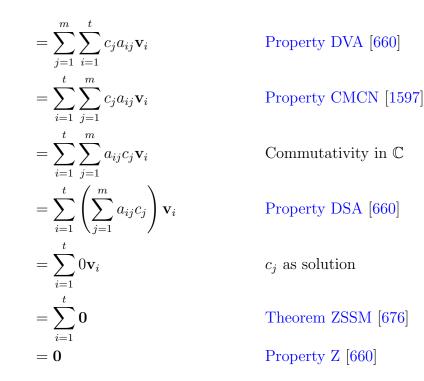
Now we form, unmotivated, the homogeneous system of t equations in the m variables, x_j , $1 \le j \le m$, where the coefficients are the just-discovered scalars a_{ij} ,

$$\sum_{j=1}^{m} a_{ij} x_j = 0 \qquad \qquad 1 \le i \le t$$

This is a homogeneous system with more variables than equations (our hypothesis is expressed as m > t), so by Theorem HMVEI [151] there are infinitely many solutions. Choose one of these solutions that is not trivial and denote it by $x_j = c_j$, $1 \le j \le m$. As a solution to the homogeneous system, we then have $\sum_{j=1}^{m} a_{ij}c_j = 0$ for $1 \le i \le t$. As a collection of nontrivial scalars, c_j , $1 \le j \le m$, will provide the nontrivial relation of linear dependence we desire,

$$\sum_{j=1}^{m} c_j \mathbf{u}_j = \sum_{j=1}^{m} c_j \left(\sum_{i=1}^{t} a_{ij} \mathbf{v}_i \right)$$

Definition TSVS [746]



That does it. R has been undeniably shown to be a linearly dependent set.

Notice how the swap of the two summations is so much easier in the third step above, as opposed to all the rearranging and regrouping that takes place in the previous proof. In about half the space. And there are no ellipses (\ldots) .

Theorem SSLD [816] can be viewed as a generalization of Theorem MVSLD [328]. We know that \mathbb{C}^m has a basis with m vectors in it (Theorem SUVB [779]), so it is a set of m vectors that spans \mathbb{C}^m . By Theorem SSLD [816], any set of more than m vectors from \mathbb{C}^m will be linearly dependent. But this is exactly the conclusion we have in Theorem MVSLD [328]. Maybe this is not a total shock, as the proofs of both theorems rely heavily on Theorem HMVEI [151]. The beauty of Theorem SSLD [816] is that it applies in any vector space. We illustrate the generality of this theorem, and hint at its power, in the next example.

Example LDP4 Linearly dependent set in P_4

In Example SSP4 [746] we showed that

$$S = \left\{ x - 2, \ x^2 - 4x + 4, \ x^3 - 6x^2 + 12x - 8, \ x^4 - 8x^3 + 24x^2 - 32x + 16 \right\}$$

is a spanning set for $W = \{p(x) | p \in P_4, p(2) = 0\}$. So we can apply Theorem SSLD [816] to W with t = 4. Here is a set of five vectors from W, as you may check by verifying that each is a polynomial of degree 4 or less and has x = 2 as a root,

$$T = \{p_1, p_2, p_3, p_4, p_5\} \subseteq W$$

$$p_{1} = x^{4} - 2x^{3} + 2x^{2} - 8x + 8$$

$$p_{2} = -x^{3} + 6x^{2} - 5x - 6$$

$$p_{3} = 2x^{4} - 5x^{3} + 5x^{2} - 7x + 2$$

$$p_{4} = -x^{4} + 4x^{3} - 7x^{2} + 6x$$

$$p_5 = 4x^3 - 9x^2 + 5x - 6$$

By Theorem SSLD [816] we conclude that T is linearly dependent, with no further computations.

Theorem SSLD [816] is indeed powerful, but our main purpose in proving it right now was to make sure that our definition of dimension (Definition D [815]) is well-defined. Here's the theorem.

Theorem BIS Bases have Identical Sizes

Suppose that V is a vector space with a finite basis B and a second basis C. Then B and C have the same size. \Box

Proof Suppose that C has more vectors than B. (Allowing for the possibility that C is infinite, we can replace C by a subset that has more vectors than B.) As a basis, B is a spanning set for V (Definition B [777]), so Theorem SSLD [816] says that C is linearly dependent. However, this contradicts the fact that as a basis C is linearly independent (Definition B [777]). So C must also be a finite set, with size less than, or equal to, that of B.

Suppose that B has more vectors than C. As a basis, C is a spanning set for V (Definition B [777]), so Theorem SSLD [816] says that B is linearly dependent. However, this contradicts the fact that as a basis B is linearly independent (Definition B [777]). So C cannot be strictly smaller than B.

The only possibility left for the sizes of B and C is for them to be equal.

Theorem BIS [823] tells us that if we find one finite basis in a vector space, then they all have the same size. This (finally) makes Definition D [815] unambiguous.

Subsection DVS Dimension of Vector Spaces

We can now collect the dimension of some common, and not so common, vector spaces.

Theorem DCM Dimension of \mathbb{C}^m The dimension of \mathbb{C}^m (Example VSCV [663]) is m.

Proof Theorem SUVB [779] provides a basis with m vectors.

Theorem DP Dimension of P_n The dimension of P_n (Example VSP [665]) is n + 1.

Proof Example BP [780] provides two bases with n + 1 vectors. Take your pick.

Theorem DM Dimension of M_{mn}

 \square

The dimension of M_{mn} (Example VSM [663]) is mn.

Proof Example BM [780] provides a basis with mn vectors.

Example DSM22 Dimension of a subspace of M_{22}

It should now be plausible that

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| 2a + b + 3c + 4d = 0, \ -a + 3b - 5c - 2d = 0 \right\}$$

is a subspace of the vector space M_{22} (Example VSM [663]). (It is.) To find the dimension of Z we must first find a basis, though any old basis will do.

First concentrate on the conditions relating a, b, c and d. They form a homogeneous system of two equations in four variables with coefficient matrix

$$\begin{bmatrix} 2 & 1 & 3 & 4 \\ -1 & 3 & -5 & -2 \end{bmatrix}$$

We can row-reduce this matrix to obtain

$$\begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

Rewrite the two equations represented by each row of this matrix, expressing the dependent variables (a and b) in terms of the free variables (c and d), and we obtain,

$$a = -2c - 2d$$

b = c

We can now write a typical entry of Z strictly in terms of c and d, and we can decompose the result,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} -2c - 2d & c \\ c & d \end{bmatrix} = \begin{bmatrix} -2c & c \\ c & 0 \end{bmatrix} + \begin{bmatrix} -2d & 0 \\ 0 & d \end{bmatrix} = c \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}$$

this equation says that an arbitrary matrix in Z can be written as a linear combination of the two vectors in

$$S = \left\{ \begin{bmatrix} -2 & 1\\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -2 & 0\\ 0 & 1 \end{bmatrix} \right\}$$

so we know that

$$Z = \langle S \rangle = \left\langle \left\{ \begin{bmatrix} -2 & 1\\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -2 & 0\\ 0 & 1 \end{bmatrix} \right\} \right\rangle$$

Are these two matrices (vectors) also linearly independent? Begin with a relation of linear dependence on S,

$$a_1 \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix} + a_2 \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix} = \mathcal{O}$$
$$\begin{bmatrix} -2a_1 - 2a_2 & a_1 \\ a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

From the equality of the two entries in the last row, we conclude that $a_1 = 0$, $a_2 = 0$. Thus the only possible relation of linear dependence is the trivial one, and therefore S is linearly independent (Definition LI [735]). So S is a basis for V (Definition B [777]). Finally, we can conclude that dim (Z) = 2 (Definition D [815]) since S has two elements.

Example DSP4 Dimension of a subspace of P_4

In Example BSP4 [781] we showed that

 $S = \left\{ x - 2, \ x^2 - 4x + 4, \ x^3 - 6x^2 + 12x - 8, \ x^4 - 8x^3 + 24x^2 - 32x + 16 \right\}$

is a basis for $W = \{ p(x) | p \in P_4, p(2) = 0 \}$. Thus, the dimension of W is four, dim (W) = 4.

Note that dim $(P_4) = 5$ by Theorem DP [824], so W is a subspace of dimension 4 within the vector space P_4 of dimension 5, illustrating the upcoming Theorem PSSD [857].

Example DC

Dimension of the crazy vector space

In Example BC [784] we determined that the set $R = \{(1, 0), (6, 3)\}$ from the crazy vector space, C (Example CVS [670]), is a basis for C. By Definition D [815] we see that C has dimension 2, dim (C) = 2.

It is possible for a vector space to have no finite bases, in which case we say it has infinite dimension. Many of the best examples of this are vector spaces of functions, which lead to constructions like Hilbert spaces. We will focus exclusively on finite-dimensional vector spaces. OK, one infinite-dimensional example, and *then* we will focus exclusively on finite-dimensional vector spaces.

Example VSPUD Vector space of polynomials with unbounded degree Define the set P by

 $P = \{ p | p(x) \text{ is a polynomial in } x \}$

Our operations will be the same as those defined for P_n (Example VSP [665]).

With no restrictions on the possible degrees of our polynomials, any finite set that is a candidate for spanning P will come up short. We will give a proof by contradiction (Technique CD [1623]). To this end, suppose that the dimension of P is finite, say dim (P) = n.

The set $T = \{1, x, x^2, ..., x^n\}$ is a linearly independent set (check this!) containing n + 1 polynomials from P. However, a basis of P will be a spanning set of P containing n vectors. This situation is a contradiction of Theorem SSLD [816], so our assumption that P has finite dimension is false. Thus, we say dim $(P) = \infty$.

Subsection RNM Rank and Nullity of a Matrix

For any matrix, we have seen that we can associate several subspaces — the null space (Theorem NSMS [705]), the column space (Theorem CSMS [719]), row space (Theorem RSMS [719]) and the left null space (Theorem LNSMS [720]). As vector spaces, each of

these has a dimension, and for the null space and column space, they are important enough to warrant names.

Definition NOM Nullity Of a Matrix

Suppose that A is an $m \times n$ matrix. Then the **nullity** of A is the dimension of the null space of A, $n(A) = \dim(\mathcal{N}(A))$. (This definition contains Notation NOM.) \triangle

Definition ROM Rank Of a Matrix

Suppose that A is an $m \times n$ matrix. Then the **rank** of A is the dimension of the column space of A, $r(A) = \dim (\mathcal{C}(A))$.

(This definition contains Notation ROM.)

Example RNM Rank and nullity of a matrix

Let's compute the rank and nullity of

$$A = \begin{bmatrix} 2 & -4 & -1 & 3 & 2 & 1 & -4 \\ 1 & -2 & 0 & 0 & 4 & 0 & 1 \\ -2 & 4 & 1 & 0 & -5 & -4 & -8 \\ 1 & -2 & 1 & 1 & 6 & 1 & -3 \\ 2 & -4 & -1 & 1 & 4 & -2 & -1 \\ -1 & 2 & 3 & -1 & 6 & 3 & -1 \end{bmatrix}$$

Δ

To do this, we will first row-reduce the matrix since that will help us determine bases for the null space and column space.

From this row-equivalent matrix in reduced row-echelon form we record $D = \{1, 3, 4, 6\}$ and $F = \{2, 5, 7\}$.

For each index in D, Theorem BCS [569] creates a single basis vector. In total the basis will have 4 vectors, so the column space of A will have dimension 4 and we write r(A) = 4.

For each index in F, Theorem BNS [334] creates a single basis vector. In total the basis will have 3 vectors, so the null space of A will have dimension 3 and we write n(A) = 3.

There were no accidents or coincidences in the previous example — with the row-reduced version of a matrix in hand, the rank and nullity are easy to compute.

Theorem CRN

Computing Rank and Nullity

Suppose that A is an $m \times n$ matrix and B is a row-equivalent matrix in reduced row-echelon form with r nonzero rows. Then r(A) = r and n(A) = n - r.

Proof Theorem BCS [569] provides a basis for the column space by choosing columns of A

that correspond to the dependent variables in a description of the solutions to $\mathcal{LS}(A, \mathbf{0})$. In the analysis of B, there is one dependent variable for each leading 1, one per nonzero row, or one per pivot column. So there are r column vectors in a basis for $\mathcal{C}(A)$.

Theorem BNS [334] provide a basis for the null space by creating basis vectors of the null space of A from entries of B, one for each independent variable, one per column with out a leading 1. So there are n - r column vectors in a basis for n(A).

Every archetype (Appendix A [1637]) that involves a matrix lists its rank and nullity. You may have noticed as you studied the archetypes that the larger the column space is the smaller the null space is. A simple corollary states this trade-off succinctly. (See Technique LC [1634].)

Theorem RPNC Rank Plus Nullity is Columns

Suppose that A is an $m \times n$ matrix. Then r(A) + n(A) = n.

Proof Let r be the number of nonzero rows in a row-equivalent matrix in reduced rowechelon form. By Theorem CRN [830],

$$r(A) + n(A) = r + (n - r) = n$$

When we first introduced r as our standard notation for the number of nonzero rows in a matrix in reduced row-echelon form you might have thought r stood for "rows." Not really — it stands for "rank"!

 \square

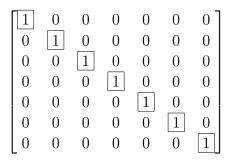
Subsection RNNM Rank and Nullity of a Nonsingular Matrix

Let's take a look at the rank and nullity of a square matrix.

Example RNSM Rank and nullity of a square matrix The matrix

$$E = \begin{bmatrix} 0 & 4 & -1 & 2 & 2 & 3 & 1 \\ 2 & -2 & 1 & -1 & 0 & -4 & -3 \\ -2 & -3 & 9 & -3 & 9 & -1 & 9 \\ -3 & -4 & 9 & 4 & -1 & 6 & -2 \\ -3 & -4 & 6 & -2 & 5 & 9 & -4 \\ 9 & -3 & 8 & -2 & -4 & 2 & 4 \\ 8 & 2 & 2 & 9 & 3 & 0 & 9 \end{bmatrix}$$

is row-equivalent to the matrix in reduced row-echelon form,



With n = 7 columns and r = 7 nonzero rows Theorem CRN [830] tells us the rank is r(E) = 7 and the nullity is n(E) = 7 - 7 = 0.

The value of either the nullity or the rank are enough to characterize a nonsingular matrix.

Theorem RNNM Rank and Nullity of a Nonsingular Matrix

Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. The rank of A is n, r(A) = n.
- 3. The nullity of A is zero, n(A) = 0.

 \square

Proof $(1 \Rightarrow 2)$ Theorem CSNM [574] says that if A is nonsingular then $\mathcal{C}(A) = \mathbb{C}^n$. If $\mathcal{C}(A) = \mathbb{C}^n$, then the column space has dimension n by Theorem DCM [824], so the rank of A is n.

 $(2 \Rightarrow 3)$ Suppose r(A) = n. Then Theorem RPNC [831] gives

n(A) = n - r(A) = n - n = 0Theorem RPNC [831]
Hypothesis

 $(3 \Rightarrow 1)$ Suppose n(A) = 0, so a basis for the null space of A is the empty set. This implies that $\mathcal{N}(A) = \{\mathbf{0}\}$ and Theorem NMTNS [179] says A is nonsingular.

With a new equivalence for a nonsingular matrix, we can update our list of equivalences (Theorem NME5 [791]) which now becomes a list requiring double digits to number.

Theorem NME6 Nonsingular Matrix Equivalences, Round 6

Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$

- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.
- 7. The column space of A is \mathbb{C}^n , $\mathcal{C}(A) = \mathbb{C}^n$.
- 8. The columns of A are a basis for \mathbb{C}^n .
- 9. The rank of A is n, r(A) = n.
- 10. The nullity of A is zero, n(A) = 0.

Proof Building on Theorem NME5 [791] we can add two of the statements from Theorem RNNM [833]. ■

Subsection READ Reading Questions

1. What is the dimension of the vector space P_6 , the set of all polynomials of degree 6 or less?

 \square

- 2. How are the rank and nullity of a matrix related?
- 3. Explain why we might say that a nonsingular matrix has "full rank."

Subsection EXC Exercises

C20 The archetypes listed below are matrices, or systems of equations with coefficient matrices. For each, compute the nullity and rank of the matrix. This information is listed for each archetype (along with the number of columns in the matrix, so as to illustrate Theorem RPNC [831]), and notice how it could have been computed immediately after the determination of the sets D and F associated with the reduced row-echelon form of the matrix.

Archetype A [1642] Archetype B [1652] Archetype C [1662] Archetype D [1670]/Archetype E [1678] Archetype F [1686] Archetype G [1697]/Archetype H [1705] Archetype I [1714] Archetype J [1723] Archetype K [1734] Archetype L [1742] Contributed by Robert Beezer

C21 Find the dimension of the subspace
$$W = \left\{ \begin{bmatrix} a+b\\a+c\\a+d\\d \end{bmatrix} \middle| a,b,c,d \in \mathbb{C} \right\}$$
 of \mathbb{C}^4 .
Contributed by Chris Black Solution [842]

Find the dimension of the subspace $W = \{a + bx + cx^2 + dx^3 | a + b + c + d = 0\}$ of C22 P_3 . Contributed by Chris Black Solution [843]

. -

Find the dimension of the subspace $W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a + b = c, b + c = d, c + d = a \right\}$ C23of $M_{2,2}$. Contributed by Chris Black Solution [843]

C30 For the matrix A below, compute the dimension of the null space of A, dim $(\mathcal{N}(A))$.

$$A = \begin{bmatrix} 2 & -1 & -3 & 11 & 9\\ 1 & 2 & 1 & -7 & -3\\ 3 & 1 & -3 & 6 & 8\\ 2 & 1 & 2 & -5 & -3 \end{bmatrix}$$

Contributed by Robert Beezer Solution [844] C31 The set W below is a subspace of \mathbb{C}^4 . Find the dimension of W.

$$W = \left\langle \left\{ \begin{bmatrix} 2\\-3\\4\\1 \end{bmatrix}, \begin{bmatrix} 3\\0\\1\\-2 \end{bmatrix}, \begin{bmatrix} -4\\-3\\2\\5 \end{bmatrix} \right\} \right\rangle$$

Contributed by Robert Beezer Solution [845]

C35 Find the rank and nullity of the matrix
$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$
.
Contributed by Chris Black Solution [846]

C36 Find the rank and nullity of the matrix
$$A = \begin{bmatrix} 1 & 2 & 1 & 1 & 1 \\ 1 & 3 & 2 & 0 & 4 \\ 1 & 2 & 1 & 1 & 1 \end{bmatrix}$$
.
Contributed by Chris Black Solution [846]

C37 Find the rank and nullity of the matrix
$$A = \begin{bmatrix} 3 & 2 & 1 & 1 & 1 \\ 2 & 3 & 0 & 1 & 1 \\ -1 & 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & -1 \end{bmatrix}$$

Contributed by Chris Black Solution [846]

C40 In Example LDP4 [822] we determined that the set of five polynomials, T, is linearly dependent by a simple invocation of Theorem SSLD [816]. Prove that T is linearly dependent from scratch, beginning with Definition LI [735]. Contributed by Robert Beezer

M20 M_{22} is the vector space of 2×2 matrices. Let S_{22} denote the set of all 2×2 symmetric matrices. That is

$$S_{22} = \left\{ A \in M_{22} | A^t = A \right\}$$

- (a) Show that S_{22} is a subspace of M_{22} .
- (b) Exhibit a basis for S_{22} and prove that it has the required properties.
- (c) What is the dimension of S_{22} ?

Contributed by Robert Beezer Solution [846]

M21 A 2 × 2 matrix *B* is upper triangular if $[B]_{21} = 0$. Let UT_2 be the set of all 2 × 2 upper triangular matrices. Then UT_2 is a subspace of the vector space of all 2 × 2 matrices, M_{22} (you may assume this). Determine the dimension of UT_2 providing *all* of the necessary justifications for your answer.

Contributed by Robert Beezer Solution [848]

Subsection SOL Solutions

C21Contributed by Chris BlackStatement [838]The subspace W can be written as

forms a basis of W. Thus, W is a subspace of \mathbb{C}^4 with dimension 4 (and must therefore equal \mathbb{C}^4).

C22 Contributed by Chris Black Statement [838] The subspace $W = \{a + bx + cx^2 + dx^3 | a + b + c + d = 0\}$ can be written as

$$W = \left\{ a + bx + cx^{2} + (-a - b - c)x^{3} \middle| a, b, c \in \mathbb{C} \right\}$$

= $\left\{ a(1 - x^{3}) + b(x - x^{3}) + c(x^{2} - x^{3}) \middle| a, b, c \in \mathbb{C} \right\}$
= $\left\langle \left\{ 1 - x^{3}, x - x^{3}, x^{2} - x^{3} \right\} \right\rangle$

Since these vectors are linearly independent (why?), W is a subspace of P_3 with dimension 3.

C23 Contributed by Chris Black Statement [838] The equations specified are equivalent to the system

$$a+b-c = 0$$
$$b+c-d = 0$$
$$a-c-d = 0$$

The coefficient matrix of this system row-reduces to

$$\begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

Thus, every solution can be decribed with a suitable choice of d, together with a = 3d, b = -d and c = 2d. Thus the subspace W can be described as

$$W = \left\{ \begin{bmatrix} 3d & -d \\ 2d & d \end{bmatrix} \middle| d \in \mathbb{C} \right\} = \left\langle \left\{ \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix} \right\} \right\rangle$$

So, W is a subspace of $M_{2,2}$ with dimension 1.

C30 Contributed by Robert Beezer Statement [838] Row reduce A,

$$A \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & -3 & -1 \\ 0 & 0 & 1 & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So r = 3 for this matrix. Then

$$\dim \left(\mathcal{N}(A) \right) = n \left(A \right)$$
$$= \left(n \left(A \right) + r \left(A \right) \right) - r \left(A \right)$$

Definition NOM [829]

$$= 5 - r (A)$$
Theorem RPNC [831]

$$= 5 - 3$$
Theorem CRN [830]

$$= 2$$

We could also use Theorem BNS [334] and create a basis for $\mathcal{N}(A)$ with n - r = 5 - 3 = 2 vectors (because the solutions are described with 2 free variables) and arrive at the dimension as the size of this basis.

C31 Contributed by Robert Beezer Statement [839] We will appeal to Theorem BS [376] (or you could consider this an appeal to Theorem BCS [569]). Put the three column vectors of this spanning set into a matrix as columns and row-reduce.

$$\begin{bmatrix} 2 & 3 & -4 \\ -3 & 0 & -3 \\ 4 & 1 & 2 \\ 1 & -2 & 5 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The pivot columns are $D = \{1, 2\}$ so we can "keep" the vectors corresponding to the pivot columns and set

$$T = \left\{ \begin{bmatrix} 2\\ -3\\ 4\\ 1 \end{bmatrix}, \begin{bmatrix} 3\\ 0\\ 1\\ -2 \end{bmatrix} \right\}$$

and conclude that $W = \langle T \rangle$ and T is linearly independent. In other words, T is a basis with

two vectors, so W has dimension 2.

C35 Contributed by Chris Black Statement [839]

with leading 1's) is 3, and the nullity is 0.

C36 Contributed by Chris Black Statement [839] The row reduced form of matrix A is $\begin{bmatrix} 1 & 0 & -1 & 3 & 5 \\ 0 & 1 & 1 & -1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, so the rank of A (number of

columns with leading 1's) is 2, and the nullity is 5-2=3.

C37 Contributed by Chris Black Statement [840] This matrix A row reduces to the 5×5 identity matrix, so it has full rank. The rank of A is 5, and the nullity is 0.

M20 Contributed by Robert Beezer Statement [840]

We will use the three criteria of Theorem TSS [698]. The zero vector of M_{22} is the zero (a) matrix, \mathcal{O} (Definition ZM [433]), which is a symmetric matrix. So S_{22} is not empty, since $\mathcal{O} \in S_{22}$.

Suppose that A and B are two matrices in S_{22} . Then we know that $A^t = A$ and $B^t = B$.

We want to know if $A + B \in S_{22}$, so test A + B for membership,

$$(A+B)^{t} = A^{t} + B^{t}$$

= A + B

$$A, B \in S_{22}$$

Theorem TMA [436]
A, B \in S_{22}

So A + B is symmetric and qualifies for membership in S_{22} . Suppose that $A \in S_{22}$ and $\alpha \in \mathbb{C}$. Is $\alpha A \in S_{22}$? We know that $A^t = A$. Now check that,

$$\alpha A^{t} = \alpha A^{t}$$
 Theorem TMSM [436]
= αA $A \in S_{22}$

So αA is also symmetric and qualifies for membership in S_{22} .

With the three criteria of Theorem TSS [698] fulfilled, we see that S_{22} is a subspace of M_{22} .

(b) An arbitrary matrix from S_{22} can be written as $\begin{bmatrix} a & b \\ b & d \end{bmatrix}$. We can express this matrix as

$$\begin{bmatrix} a & b \\ b & d \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ b & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}$$
$$= a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

this equation says that the set

$$T = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

spans S_{22} . Is it also linearly independent?

Write a relation of linear dependence on S,

$$\mathcal{O} = a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix}$$

The equality of these two matrices (Definition ME [427]) tells us that $a_1 = a_2 = a_3 = 0$, and the only relation of linear dependence on T is trivial. So T is linearly independent, and hence is a basis of S_{22} .

(c) The basis T found in part (b) has size 3. So by Definition D [815], dim $(S_{22}) = 3$.

M21Contributed by Robert BeezerStatement [840]A typical matrix from UT_2 looks like

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix}$$

where $a, b, c \in \mathbb{C}$ are arbitrary scalars. Observing this we can then write

$$\begin{bmatrix} a & b \\ 0 & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

which says that

$$R = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is a spanning set for UT_2 (Definition TSVS [746]). Is R is linearly independent? If so, it is a basis for UT_2 . So consider a relation of linear dependence on R,

$$\alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \mathcal{O} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

From this equation, one rapidly arrives at the conclusion that $\alpha_1 = \alpha_2 = \alpha_3 = 0$. So R is a linearly independent set (Definition LI [735]), and hence is a basis (Definition B [777]) for UT_2 . Now, we simply count up the size of the set R to see that the dimension of UT_2 is dim $(UT_2) = 3$.

Section PD Properties of Dimension

Once the dimension of a vector space is known, then the determination of whether or not a set of vectors is linearly independent, or if it spans the vector space, can often be much easier. In this section we will state a workhorse theorem and then apply it to the column space and row space of a matrix. It will also help us describe a super-basis for \mathbb{C}^m .

Subsection GT Goldilocks' Theorem

We begin with a useful theorem that we will need later, and in the proof of the main theorem in this subsection. This theorem says that we can extend linearly independent sets, one vector at a time, by adding vectors from outside the span of the linearly independent set, all the while preserving the linear independence of the set.

Theorem ELIS Extending Linearly Independent Sets

Suppose V is vector space and S is a linearly independent set of vectors from V. Suppose

w is a vector such that $\mathbf{w} \notin \langle S \rangle$. Then the set $S' = S \cup \{\mathbf{w}\}$ is linearly independent.

Proof Suppose $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_m}$ and begin with a relation of linear dependence on S',

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \dots + a_m\mathbf{v}_m + a_{m+1}\mathbf{w} = \mathbf{0}.$$

There are two cases to consider. First suppose that $a_{m+1} = 0$. Then the relation of linear dependence on S' becomes

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \dots + a_m\mathbf{v}_m = \mathbf{0}.$$

and by the linear independence of the set S, we conclude that $a_1 = a_2 = a_3 = \cdots = a_m = 0$. So all of the scalars in the relation of linear dependence on S' are zero.

In the second case, suppose that $a_{m+1} \neq 0$. Then the relation of linear dependence on S' becomes

$$a_{m+1}\mathbf{w} = -a_1\mathbf{v}_1 - a_2\mathbf{v}_2 - a_3\mathbf{v}_3 - \dots - a_m\mathbf{v}_m$$
$$\mathbf{w} = -\frac{a_1}{a_{m+1}}\mathbf{v}_1 - \frac{a_2}{a_{m+1}}\mathbf{v}_2 - \frac{a_3}{a_{m+1}}\mathbf{v}_3 - \dots - \frac{a_m}{a_{m+1}}\mathbf{v}_m$$

(

This equation expresses \mathbf{w} as a linear combination of the vectors in S, contrary to the assumption that $\mathbf{w} \notin \langle S \rangle$, so this case leads to a contradiction.

The first case yielded only a trivial relation of linear dependence on S' and the second case led to a contradiction. So S' is a linearly independent set since any relation of linear dependence is trivial.

In the story *Goldilocks and the Three Bears*, the young girl Goldilocks visits the empty house of the three bears while out walking in the woods. One bowl of porridge is too hot,

 \square

the other too cold, the third is just right. One chair is too hard, one too soft, the third is just right. So it is with sets of vectors — some are too big (linearly dependent), some are too small (they don't span), and some are just right (bases). Here's Goldilocks' Theorem.

Theorem G

Goldilocks

Suppose that V is a vector space of dimension t. Let $S = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_m}$ be a set of vectors from V. Then

- 1. If m > t, then S is linearly dependent.
- 2. If m < t, then S does not span V.

3. If m = t and S is linearly independent, then S spans V.

4. If m = t and S spans V, then S is linearly independent.

Proof Let *B* be a basis of *V*. Since dim (V) = t, Definition B [777] and Theorem BIS [823] imply that *B* is a linearly independent set of *t* vectors that spans *V*.

- 1. Suppose to the contrary that S is linearly independent. Then B is a smaller set of vectors that spans V. This contradicts Theorem SSLD [816].
- 2. Suppose to the contrary that S does span V. Then B is a larger set of vectors that is linearly independent. This contradicts Theorem SSLD [816].

 \square

- 3. Suppose to the contrary that S does not span V. Then we can choose a vector \mathbf{w} such that $\mathbf{w} \in V$ and $\mathbf{w} \notin \langle S \rangle$. By Theorem ELIS [850], the set $S' = S \cup \{\mathbf{w}\}$ is again linearly independent. Then S' is a set of m + 1 = t + 1 vectors that are linearly independent, while B is a set of t vectors that span V. This contradicts Theorem SSLD [816].
- 4. Suppose to the contrary that S is linearly dependent. Then by Theorem DLDS [365] (which can be upgraded, with no changes in the proof, to the setting of a general vector space), there is a vector in S, say \mathbf{v}_k that is equal to a linear combination of the other vectors in S. Let $S' = S \setminus {\{\mathbf{v}_k\}}$, the set of "other" vectors in S. Then it is easy to show that $V = \langle S \rangle = \langle S' \rangle$. So S' is a set of m 1 = t 1 vectors that spans V, while B is a set of t linearly independent vectors in V. This contradicts Theorem SSLD [816].

There is a tension in the construction of basis. Make a set too big and you will end up with relations of linear dependence among the vectors. Make a set too small and you will not have enough raw material to span the entire vector space. Make a set just the right size (the dimension) and you only need to have linear independence or spanning, and you get the other property for free. These roughly-stated ideas are made precise by Theorem G [852].

The structure and proof of this theorem also deserve comment. The hypotheses seem innocuous. We presume we know the dimension of the vector space in hand, then we mostly just look at the size of the set S. From this we get big conclusions about spanning and linear independence. Each of the four proofs relies on ultimately contradicting Theorem SSLD [816], so in a way we could think of this entire theorem as a corollary of Theorem SSLD

[816]. (See Technique LC [1634].) The proofs of the third and fourth parts parallel each other in style (add \mathbf{w} , toss \mathbf{v}_k) and then turn on Theorem ELIS [850] before contradicting Theorem SSLD [816].

Theorem G [852] is useful in both concrete examples and as a tool in other proofs. We will use it often to bypass verifying linear independence or spanning.

Example BPR Bases for P_n , reprised

In Example BP [780] we claimed that

$$B = \{1, x, x^2, x^3, \dots, x^n\}$$

$$C = \{1, 1+x, 1+x+x^2, 1+x+x^2+x^3, \dots, 1+x+x^2+x^3+\dots+x^n\}$$

were both bases for P_n (Example VSP [665]). Suppose we had first verified that B was a basis, so we would then know that dim $(P_n) = n + 1$. The size of C is n + 1, the right size to be a basis. We could then verify that C is linearly independent. We would not have to make any special efforts to prove that C spans P_n , since Theorem G [852] would allow us to conclude this property of C directly. Then we would be able to say that C is a basis of P_n also.

Example BDM22 Basis by dimension in M_{22} In Example DSM22 [825] we showed that

$$B = \left\{ \begin{bmatrix} -2 & 1\\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -2 & 0\\ 0 & 1 \end{bmatrix} \right\}$$

is a basis for the subspace Z of M_{22} (Example VSM [663]) given by

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| 2a + b + 3c + 4d = 0, \ -a + 3b - 5c - d = 0 \right\}$$

This tells us that dim (Z) = 2. In this example we will find another basis. We can construct two new matrices in Z by forming linear combinations of the matrices in B.

$$2\begin{bmatrix} -2 & 1\\ 1 & 0 \end{bmatrix} + (-3)\begin{bmatrix} -2 & 0\\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2\\ 2 & -3 \end{bmatrix}$$
$$3\begin{bmatrix} -2 & 1\\ 1 & 0 \end{bmatrix} + 1\begin{bmatrix} -2 & 0\\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -8 & 3\\ 3 & 1 \end{bmatrix}$$

Then the set

$$C = \left\{ \begin{bmatrix} 2 & 2 \\ 2 & -3 \end{bmatrix}, \begin{bmatrix} -8 & 3 \\ 3 & 1 \end{bmatrix} \right\}$$

has the right size to be a basis of Z. Let's see if it is a linearly independent set. The relation of linear dependence

$$a_{1}\begin{bmatrix} 2 & 2\\ 2 & -3 \end{bmatrix} + a_{2}\begin{bmatrix} -8 & 3\\ 3 & 1 \end{bmatrix} = \mathcal{O}$$
$$\begin{bmatrix} 2a_{1} - 8a_{2} & 2a_{1} + 3a_{2}\\ 2a_{1} + 3a_{2} & -3a_{1} + a_{2} \end{bmatrix} = \begin{bmatrix} 0 & 0\\ 0 & 0 \end{bmatrix}$$

leads to the homogeneous system of equations whose coefficient matrix

$$\begin{bmatrix} 2 & -8 \\ 2 & 3 \\ 2 & 3 \\ -3 & 1 \end{bmatrix}$$

row-reduces to

So with $a_1 = a_2 = 0$ as the only solution, the set is linearly independent. Now we can apply Theorem G [852] to see that C also spans Z and therefore is a second basis for Z.

 $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$

Example SVP4 Sets of vectors in P_4 In Example PSP4 [781] we showed the

In Example BSP4 [781] we showed that

$$B = \left\{ x - 2, \ x^2 - 4x + 4, \ x^3 - 6x^2 + 12x - 8, \ x^4 - 8x^3 + 24x^2 - 32x + 16 \right\}$$

is a basis for $W = \{ p(x) | p \in P_4, p(2) = 0 \}$. So dim (W) = 4.

The set

{
$$3x^2 - 5x - 2, 2x^2 - 7x + 6, x^3 - 2x^2 + x - 2$$
}

is a subset of W (check this) and it happens to be linearly independent (check this, too). However, by Theorem G [852] it cannot span W.

The set

$$\left\{3x^2-5x-2, 2x^2-7x+6, x^3-2x^2+x-2, -x^4+2x^3+5x^2-10x, x^4-16\right\}$$

is another subset of W (check this) and Theorem G [852] tells us that it must be linearly dependent.

The set

$$\left\{x-2, x^2-2x, x^3-2x^2, x^4-2x^3\right\}$$

is a third subset of W (check this) and is linearly independent (check this). Since it has the right size to be a basis, and is linearly independent, Theorem G [852] tells us that it also spans W, and therefore is a basis of W.

A simple consequence of Theorem G [852] is the observation that proper subspaces have strictly smaller dimensions. Hopefully this may seem intuitively obvious, but it still requires proof, and we will cite this result later.

Theorem PSSD

Proper Subspaces have Smaller Dimension

Suppose that U and V are subspaces of the vector space W, such that $U \subsetneq V$. Then $\dim(U) < \dim(V)$.

Proof Suppose that dim (U) = m and dim (V) = t. Then U has a basis B of size m. If m > t, then by Theorem G [852], B is linearly dependent, which is a contradiction. If m = t,

 \boxtimes

then by Theorem G [852], B spans V. Then $U = \langle B \rangle = V$, also a contradiction. All that remains is that m < t, which is the desired conclusion.

The final theorem of this subsection is an extremely powerful tool for establishing the equality of two sets that are subspaces. Notice that the hypotheses include the equality of two integers (dimensions) while the conclusion is the equality of two sets (subspaces). It is the extra "structure" of a vector space and its dimension that makes possible this huge leap from an integer equality to a set equality.

Theorem EDYES

Equal Dimensions Yields Equal Subspaces

Suppose that U and V are subspaces of the vector space W, such that $U \subseteq V$ and dim $(U) = \dim(V)$. Then U = V.

Proof We give a proof by contradiction (Technique CD [1623]). Suppose to the contrary that $U \neq V$. Since $U \subseteq V$, there must be a vector \mathbf{v} such that $\mathbf{v} \in V$ and $\mathbf{v} \notin U$. Let $B = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t}$ be a basis for U. Then, by Theorem ELIS [850], the set $C = B \cup {\mathbf{v}} = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t, \mathbf{v}}$ is a linearly independent set of t + 1 vectors in V. However, by hypothesis, V has the same dimension as U (namely t) and therefore Theorem G [852] says that C is too big to be linearly independent. This contradiction shows that U = V.

Subsection RT Ranks and Transposes

We now prove one of the most surprising theorems about matrices. Notice the paucity of hypotheses compared to the precision of the conclusion.

Theorem RMRT Rank of a Matrix is the Rank of the Transpose Suppose A is an $m \times n$ matrix. Then $r(A) = r(A^t)$.

Proof Suppose we row-reduce A to the matrix B in reduced row-echelon form, and B has r non-zero rows. The quantity r tells us three things about B: the number of leading 1's, the number of non-zero rows and the number of pivot columns. For this proof we will be interested in the latter two.

Theorem BRS [583] and Theorem BCS [569] each has a conclusion that provides a basis, for the row space and the column space, respectively. In each case, these bases contain r vectors. This observation makes the following go.

$r(A) = \dim \left(\mathcal{C}(A) \right)$	Definition ROM [829]
= r	Theorem BCS [569]
$= \dim \left(\mathcal{R}(A) \right)$	Theorem BRS [583]

 \square

 $= \dim (\mathcal{C}(A^{t}))$ Theorem CSRST [586] $= r (A^{t})$ Definition ROM [829]

Jacob Linenthal helped with this proof.

This says that the row space and the column space of a matrix have the same dimension, which should be very surprising. It does *not* say that column space and the row space are identical. Indeed, if the matrix is not square, then the sizes (number of slots) of the vectors in each space are different, so the sets are not even comparable.

It is not hard to construct by yourself examples of matrices that illustrate Theorem RMRT [859], since it applies equally well to *any* matrix. Grab a matrix, row-reduce it, count the nonzero rows or the leading 1's. That's the rank. Transpose the matrix, row-reduce that, count the nonzero rows or the leading 1's. That's the rank of the transpose. The theorem says the two will be equal. Here's an example anyway.

Example RRTI Rank, rank of transpose, Archetype I

Archetype I [1714] has a 4×7 coefficient matrix which row-reduces to

$$\begin{bmatrix} 1 & 4 & 0 & 0 & 2 & 1 & -3 \\ 0 & 0 & 1 & 0 & 1 & -3 & 5 \\ 0 & 0 & 0 & 1 & 2 & -6 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

so the rank is 3. Row-reducing the transpose yields

$$\begin{bmatrix}
1 & 0 & 0 & -\frac{31}{7} \\
0 & 1 & 0 & \frac{12}{7} \\
0 & 0 & 1 & \frac{13}{7} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}$$

demonstrating that the rank of the transpose is also 3.

Subsection DFS Dimension of Four Subspaces

That the rank of a matrix equals the rank of its transpose is a fundamental and surprising result. However, applying Theorem FS [625] we can easily determine the dimension of all four fundamental subspaces associated with a matrix.

Theorem DFS Dimensions of Four Subspaces

Suppose that A is an $m \times n$ matrix, and B is a row-equivalent matrix in reduced row-echelon

 \boxtimes

form with r nonzero rows. Then

- 1. dim $(\mathcal{N}(A)) = n r$
- 2. dim $(\mathcal{C}(A)) = r$
- 3. dim $(\mathcal{R}(A)) = r$
- 4. dim $(\mathcal{L}(A)) = m r$

Proof If A row-reduces to a matrix in reduced row-echelon form with r nonzero rows, then the matrix C of extended echelon form (Definition EEF [619]) will be an $r \times n$ matrix in reduced row-echelon form with no zero rows and r pivot columns (Theorem PEEF [621]). Similarly, the matrix L of extended echelon form (Definition EEF [619]) will be an $m - r \times m$ matrix in reduced row-echelon form with no zero rows and m - r pivot columns (Theorem PEEF [621]).

$$\dim (\mathcal{N}(A)) = \dim (\mathcal{N}(C)) \qquad \text{Theorem FS [625]} \\ = n - r \qquad \text{Theorem BNS [334]} \\ \dim (\mathcal{C}(A)) = \dim (\mathcal{N}(L)) \qquad \text{Theorem FS [625]} \\ = m - (m - r) \qquad \text{Theorem BNS [334]} \\ = r \end{aligned}$$

 \square

$\dim\left(\mathcal{R}(A)\right) = \dim\left(\mathcal{R}(C)\right)$	Theorem FS $[625]$
= r	Theorem BRS [583]
$\dim\left(\mathcal{L}(A)\right) = \dim\left(\mathcal{R}(L)\right)$	Theorem FS [625]
= m - r	Theorem BRS [583]

There are many different ways to state and prove this result, and indeed, the equality of the dimensions of the column space and row space is just a slight expansion of Theorem RMRT [859]. However, we have restricted our techniques to applying Theorem FS [625] and then determining dimensions with bases provided by Theorem BNS [334] and Theorem BRS [583]. This provides an appealing symmetry to the results and the proof.

Subsection DS Direct Sums

Some of the more advanced ideas in linear algebra are closely related to decomposing (Technique DC [1628]) vector spaces into direct sums of subspaces. With our previous results

about bases and dimension, now is the right time to state and collect a few results about direct sums, though we will only mention these results in passing until we get to Section NLT [1435], where they will get a heavy workout.

A direct sum is a short-hand way to describe the relationship between a vector space and two, or more, of its subspaces. As we will use it, it is not a way to construct new vector spaces from others.

Definition DS

Direct Sum

Suppose that V is a vector space with two subspaces U and W such that for every $\mathbf{v} \in V$,

1. There exists vectors $\mathbf{u} \in U$, $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$

2. If
$$\mathbf{v} = \mathbf{u}_1 + \mathbf{w}_1$$
 and $\mathbf{v} = \mathbf{u}_2 + \mathbf{w}_2$ where $\mathbf{u}_1, \mathbf{u}_2 \in U, \mathbf{w}_1, \mathbf{w}_2 \in W$ then $\mathbf{u}_1 = \mathbf{u}_2$ and $\mathbf{w}_1 = \mathbf{w}_2$.

Then V is the **direct sum** of U and W and we write $V = U \oplus W$. (This definition contains Notation DS.)

Informally, when we say V is the direct sum of the subspaces U and W, we are saying that each vector of V can always be expressed as the sum of a vector from U and a vector from W, and this expression can only be accomplished in one way (i.e. uniquely). This statement should begin to feel something like our definitions of nonsingular matrices (Definition NM [173]) and linear independence (Definition LI [735]). It should not be hard to imagine the natural extension of this definition to the case of more than two subspaces. Could you provide a careful definition of $V = U_1 \oplus U_2 \oplus U_3 \oplus \ldots \oplus U_m$ (Exercise PD.M50 [876])?

 \triangle

Example SDS Simple direct sum In \mathbb{C}^3 , define

$$\mathbf{v}_1 = \begin{bmatrix} 3\\2\\5 \end{bmatrix} \qquad \mathbf{v}_2 = \begin{bmatrix} -1\\2\\1 \end{bmatrix} \qquad \mathbf{v}_3 = \begin{bmatrix} 2\\1\\-2 \end{bmatrix}$$

Then $\mathbb{C}^3 = \langle \{\mathbf{v}_1, \mathbf{v}_2\} \rangle \oplus \langle \{\mathbf{v}_3\} \rangle$. This statement derives from the fact that $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is basis for \mathbb{C}^3 . The spanning property of B yields the decomposition of any vector into a sum of vectors from the two subspaces, and the linear independence of B yields the uniqueness of the decomposition. We will illustrate these claims with a numerical example.

Choose
$$\mathbf{v} = \begin{bmatrix} 10\\1\\6 \end{bmatrix}$$
. Then
 $\mathbf{v} = 2\mathbf{v}_1 + (-2)\mathbf{v}_2 + 1\mathbf{v}_3 = (2\mathbf{v}_1 + (-2)\mathbf{v}_2) + (1\mathbf{v}_3)$

where we have added parentheses for emphasis. Obviously $1\mathbf{v}_3 \in \langle \{\mathbf{v}_3\} \rangle$, while $2\mathbf{v}_1 + (-2)\mathbf{v}_2 \in \langle \{\mathbf{v}_1, \mathbf{v}_2\} \rangle$. Theorem VRRB [755] provides the uniqueness of the scalars in these linear combinations.

Example SDS [865] is easy to generalize into a theorem.

Theorem DSFB Direct Sum From a Basis Suppose that V is a vector space with a basis $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n}$ and $m \leq n$. Define

$$U = \langle \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \dots, \, \mathbf{v}_m\} \rangle \qquad \qquad W = \langle \{\mathbf{v}_{m+1}, \, \mathbf{v}_{m+2}, \, \mathbf{v}_{m+3}, \, \dots, \, \mathbf{v}_n\} \rangle$$

Then $V = U \oplus W$.

Proof Choose any vector $\mathbf{v} \in V$. Then by Theorem VRRB [755] there are unique scalars, $a_1, a_2, a_3, \ldots, a_n$ such that

$$\mathbf{v} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_n \mathbf{v}_n$$

= $(a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_m \mathbf{v}_m) +$
 $(a_{m+1} \mathbf{v}_{m+1} + a_{m+2} \mathbf{v}_{m+2} + a_{m+3} \mathbf{v}_{m+3} + \dots + a_n \mathbf{v}_n)$
= $\mathbf{u} + \mathbf{w}$

where we have implicitly defined **u** and **w** in the last line. It should be clear that $\mathbf{u} \in U$, and similarly, $\mathbf{w} \in W$ (and not simply by the choice of their names).

Suppose we had another decomposition of \mathbf{v} , say $\mathbf{v} = \mathbf{u}^* + \mathbf{w}^*$. Then we could write \mathbf{u}^* as a linear combination of \mathbf{v}_1 through \mathbf{v}_m , say using scalars $b_1, b_2, b_3, \ldots, b_m$. And we could write \mathbf{w}^* as a linear combination of \mathbf{v}_{m+1} through \mathbf{v}_n , say using scalars $c_1, c_2, c_3, \ldots, c_{n-m}$. These two collections of scalars would then together give a linear combination of \mathbf{v}_1 through \mathbf{v}_n that equals \mathbf{v} . By the uniqueness of $a_1, a_2, a_3, \ldots, a_n, a_i = b_i$ for $1 \le i \le m$ and $a_{m+i} = c_i$ for $1 \le i \le n-m$. From the equality of these scalars we conclude that $\mathbf{u} = \mathbf{u}^*$ and $\mathbf{w} = \mathbf{w}^*$. So with both conditions of Definition DS [864] fulfilled we see that $V = U \oplus W$.

Given one subspace of a vector space, we can always find another subspace that will pair with the first to form a direct sum. The main idea of this theorem, and its proof, is the idea of extending a linearly independent subset into a basis with repeated applications of Theorem ELIS [850].

Theorem DSFOS

Direct Sum From One Subspace

Suppose that U is a subspace of the vector space V. Then there exists a subspace W of V such that $V = U \oplus W$.

Proof If U = V, then choose $W = \{0\}$. Otherwise, choose a basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m\}$ for U. Then since B is a linearly independent set, Theorem ELIS [850] tells us there is a vector \mathbf{v}_{m+1} in V, but not in U, such that $B \cup \{\mathbf{v}_{m+1}\}$ is linearly independent. Define the subspace $U_1 = \langle B \cup \{\mathbf{v}_{m+1}\} \rangle$.

We can repeat this procedure, in the case were $U_1 \neq V$, creating a new vector \mathbf{v}_{m+2} in V, but not in U_1 , and a new subspace $U_2 = \langle B \cup \{\mathbf{v}_{m+1}, \mathbf{v}_{m+2}\}\rangle$. If we continue repeating this procedure, eventually, $U_k = V$ for some k, and we can no longer apply Theorem ELIS [850]. No matter, in this case $B \cup \{\mathbf{v}_{m+1}, \mathbf{v}_{m+2}, \ldots, \mathbf{v}_{m+k}\}$ is a linearly independent set that spans V, i.e. a basis for V.

Define $W = \langle \{\mathbf{v}_{m+1}, \mathbf{v}_{m+2}, \dots, \mathbf{v}_{m+k}\} \rangle$. We now are exactly in position to apply Theorem DSFB [865] and see that $V = U \oplus W$.

There are several different ways to define a direct sum. Our next two theorems give equivalences (Technique E [1619]) for direct sums, and therefore could have been employed as definitions. The first should further cement the notion that a direct sum has some connection with linear independence.

Theorem DSZV

Direct Sums and Zero Vectors

Suppose U and W are subspaces of the vector space V. Then $V = U \oplus W$ if and only if

1. For every $\mathbf{v} \in V$, there exists vectors $\mathbf{u} \in U$, $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$.

2. Whenever $\mathbf{0} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$, $\mathbf{w} \in W$ then $\mathbf{u} = \mathbf{w} = \mathbf{0}$.

Proof The first condition is identical in the definition and the theorem, so we only need to establish the equivalence of the second conditions.

(⇒) Assume that $V = U \oplus W$, according to Definition DS [864]. By Property Z [660], $\mathbf{0} \in V$ and $\mathbf{0} = \mathbf{0} + \mathbf{0}$. If we also assume that $\mathbf{0} = \mathbf{u} + \mathbf{w}$, then the uniqueness of the decomposition gives $\mathbf{u} = \mathbf{0}$ and $\mathbf{w} = \mathbf{0}$.

(\Leftarrow) Suppose that $\mathbf{v} \in V$, $\mathbf{v} = \mathbf{u}_1 + \mathbf{w}_1$ and $\mathbf{v} = \mathbf{u}_2 + \mathbf{w}_2$ where \mathbf{u}_1 , $\mathbf{u}_2 \in U$, \mathbf{w}_1 , $\mathbf{w}_2 \in W$. Then

 $0 = \mathbf{v} - \mathbf{v}$ $= (\mathbf{u}_1 + \mathbf{w}_1) - (\mathbf{u}_2 + \mathbf{w}_2)$ $= (\mathbf{u}_1 - \mathbf{u}_2) + (\mathbf{w}_1 - \mathbf{w}_2)$ Property AA [660]

By Property AC [660], $\mathbf{u}_1 - \mathbf{u}_2 \in U$ and $\mathbf{w}_1 - \mathbf{w}_2 \in W$. We can now apply our hypothesis, the second statement of the theorem, to conclude that

$$u_1 - u_2 = 0$$
 $w_1 - w_2 = 0$

$$\mathbf{u}_1 = \mathbf{u}_2$$
 $\mathbf{w}_1 = \mathbf{w}_2$

which establishes the uniqueness needed for the second condition of the definition.

Our second equivalence lends further credence to calling a direct sum a decomposition. The two subspaces of a direct sum have no (nontrivial) elements in common.

Theorem DSZI Direct Sums and Zero Intersection

Suppose U and W are subspaces of the vector space V. Then $V = U \oplus W$ if and only if

- 1. For every $\mathbf{v} \in V$, there exists vectors $\mathbf{u} \in U$, $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$.
- 2. $U \cap W = \{0\}.$

Proof The first condition is identical in the definition and the theorem, so we only need to establish the equivalence of the second conditions.

 (\Rightarrow) Assume that $V = U \oplus W$, according to Definition DS [864]. By Property Z [660] and Definition SI [1608], $\{0\} \subseteq U \cap W$. To establish the opposite inclusion, suppose that $\mathbf{x} \in U \cap W$. Then, since \mathbf{x} is an element of both U and W, we can write two decompositions of \mathbf{x} as a vector from U plus a vector from W,

$$\mathbf{x} = \mathbf{x} + \mathbf{0}$$
 $\mathbf{x} = \mathbf{0} + \mathbf{x}$

 \square

By the uniqueness of the decomposition, we see (twice) that $\mathbf{x} = \mathbf{0}$ and $U \cap W \subseteq \{\mathbf{0}\}$. Applying Definition SE [1604], we have $U \cap W = \{\mathbf{0}\}$.

(\Leftarrow) Assume that $U \cap W = \{\mathbf{0}\}$. And assume further that $\mathbf{v} \in V$ is such that $\mathbf{v} = \mathbf{u}_1 + \mathbf{w}_1$ and $\mathbf{v} = \mathbf{u}_2 + \mathbf{w}_2$ where $\mathbf{u}_1, \mathbf{u}_2 \in U, \mathbf{w}_1, \mathbf{w}_2 \in W$. Define $\mathbf{x} = \mathbf{u}_1 - \mathbf{u}_2$. then by Property AC [660], $\mathbf{x} \in U$. Also

$$\mathbf{x} = \mathbf{u}_1 - \mathbf{u}_2$$

= $(\mathbf{v} - \mathbf{w}_1) - (\mathbf{v} - \mathbf{w}_2)$
= $(\mathbf{v} - \mathbf{v}) - (\mathbf{w}_1 - \mathbf{w}_2)$
= $\mathbf{w}_2 - \mathbf{w}_1$

So $\mathbf{x} \in W$ by Property AC [660]. Thus, $\mathbf{x} \in U \cap W = \{\mathbf{0}\}$ (Definition SI [1608]). So $\mathbf{x} = \mathbf{0}$ and

$$\begin{aligned} \mathbf{u}_1 - \mathbf{u}_2 &= \mathbf{0} & \mathbf{w}_2 - \mathbf{w}_1 &= \mathbf{0} \\ \mathbf{u}_1 &= \mathbf{u}_2 & \mathbf{w}_2 &= \mathbf{w}_1 \end{aligned}$$

yielding the desired uniqueness of the second condition of the definition.

If the statement of Theorem DSZV [868] did not remind you of linear independence, the next theorem should establish the connection.

Theorem DSLI Direct Sums and Linear Independence

Suppose U and W are subspaces of the vector space V with $V = U \oplus W$. Suppose that R is

a linearly independent subset of U and S is a linearly independent subset of W. Then $R \cup S$ is a linearly independent subset of V.

Proof Let $R = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k}$ and $S = {\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_\ell}$. Begin with a relation of linear dependence (Definition RLD [735]) on the set $R \cup S$ using scalars $a_1, a_2, a_3, \ldots, a_k$ and $b_1, b_2, b_3, \ldots, b_\ell$. Then,

$$0 = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_k \mathbf{u}_k + b_1 \mathbf{w}_1 + b_2 \mathbf{w}_2 + b_3 \mathbf{w}_3 + \dots + b_\ell \mathbf{w}_\ell$$

= $(a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_k \mathbf{u}_k) + (b_1 \mathbf{w}_1 + b_2 \mathbf{w}_2 + b_3 \mathbf{w}_3 + \dots + b_\ell \mathbf{w}_\ell)$
= $\mathbf{u} + \mathbf{w}$

where we have made an implicit definition of the vectors $\mathbf{u} \in U$, $\mathbf{w} \in W$. Applying Theorem DSZV [868] we conclude that

$$\mathbf{u} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_k \mathbf{u}_k = \mathbf{0}$$

$$\mathbf{w} = b_1 \mathbf{w}_1 + b_2 \mathbf{w}_2 + b_3 \mathbf{w}_3 + \dots + b_\ell \mathbf{w}_\ell = \mathbf{0}$$

Now the linear independence of R and S (individually) yields

$$a_1 = a_2 = a_3 = \dots = a_k = 0$$
 $b_1 = b_2 = b_3 = \dots = b_\ell = 0$

Forced to acknowledge that only a trivial linear combination yields the zero vector, Definition LI [735] says the set $R \cup S$ is linearly independent in V.

Our last theorem in this collection will go some ways towards explaining the word "sum" in the moniker "direct sum," while also partially explaining why these results appear in a section devoted to a discussion of dimension.

Theorem DSD Direct Sums and Dimension

Suppose U and W are subspaces of the vector space V with $V = U \oplus W$. Then dim $(V) = \dim(U) + \dim(W)$.

Proof We will establish this equality of positive integers with two inequalities. We will need a basis of U (call it B) and a basis of W (call it C).

First, note that B and C have sizes equal to the dimensions of the respective subspaces. The union of these two linearly independent sets, $B \cup C$ will be linearly independent in V by Theorem DSLI [870]. Further, the two bases have no vectors in common by Theorem DSZI [869], since $B \cap C \subseteq \{0\}$ and the zero vector is never an element of a linearly independent set (Exercise LI.T10 [346]). So the size of the union is exactly the sum of the dimensions of U and W. By Theorem G [852] the size of $B \cup C$ cannot exceed the dimension of V without being linearly dependent. These observations give us dim $(U) + \dim (W) \leq \dim (V)$.

Grab any vector $\mathbf{v} \in V$. Then by Theorem DSZI [869] we can write $\mathbf{v} = \mathbf{u} + \mathbf{w}$ with $\mathbf{u} \in U$ and $\mathbf{w} \in W$. Individually, we can write \mathbf{u} as a linear combination of the basis elements in B, and similarly, we can write \mathbf{w} as a linear combination of the basis elements in C, since the bases are spanning sets for their respective subspaces. These two sets of scalars will provide a linear combination of all of the vectors in $B \cup C$ which will equal \mathbf{v} . The upshot of this is that $B \cup C$ is a spanning set for V. By Theorem G [852], the size of $B \cup C$ cannot be smaller than the dimension of V without failing to span V. These observations give us $\dim(U) + \dim(W) \ge \dim(V)$.

There is a certain appealling symmetry in the previous proof, where both linear inde-

pendence and spanning properties of the bases are used, both of the first two conclusions of Theorem G [852] are employed, and we have quoted both of the two conditions of Theorem DSZI [869].

One final theorem tells us that we can successively decompose direct sums into sums of smaller and smaller subspaces.

Theorem RDS

Repeated Direct Sums

Suppose V is a vector space with subspaces U and W with $V = U \oplus W$. Suppose that X and Y are subspaces of W with $W = X \oplus Y$. Then $V = U \oplus X \oplus Y$.

Proof Suppose that $\mathbf{v} \in V$. Then due to $V = U \oplus W$, there exist vectors $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. Due to $W = X \oplus Y$, there exist vectors $\mathbf{x} \in X$ and $\mathbf{y} \in Y$ such that $\mathbf{w} = \mathbf{x} + \mathbf{y}$. All together,

$$\mathbf{v} = \mathbf{u} + \mathbf{w} = \mathbf{u} + \mathbf{x} + \mathbf{y}$$

which would be the first condition of a definition of a 3-way direct product. Now consider the uniqueness. Suppose that

$$\mathbf{v} = \mathbf{u}_1 + \mathbf{x}_1 + \mathbf{y}_1 \qquad \qquad \mathbf{v} = \mathbf{u}_2 + \mathbf{x}_2 + \mathbf{y}_2$$

Because $\mathbf{x}_1 + \mathbf{y}_1 \in W$, $\mathbf{x}_2 + \mathbf{y}_2 \in W$, and $V = U \oplus W$, we conclude that

$$\mathbf{u}_1 = \mathbf{u}_2 \qquad \qquad \mathbf{x}_1 + \mathbf{y}_1 = \mathbf{x}_2 + \mathbf{y}_2$$

From the second equality, an application of $W = X \oplus Y$ yields the conclusions $\mathbf{x}_1 = \mathbf{x}_2$ and $\mathbf{y}_1 = \mathbf{y}_2$. This establishes the uniqueness of the decomposition of \mathbf{v} into a sum of vectors from U, X and Y.

Remember that when we write $V = U \oplus W$ there always needs to be a "superspace," in this case V. The statement $U \oplus W$ is meaningless. Writing $V = U \oplus W$ is simply a shorthand for a somewhat complicated relationship between V, U and W, as described in the two conditions of Definition DS [864], or Theorem DSZV [868], or Theorem DSZI [869]. Theorem DSFB [865] and Theorem DSFOS [867] gives us sure-fire ways to build direct sums, while Theorem DSLI [870], Theorem DSD [872] and Theorem RDS [873] tell us interesting properties of direct sums. This subsection has been long on theorems and short on examples. If we were to use the term "lemma" we might have chosen to label some of these results as such, since they will be important tools in other proofs, but may not have much interest on their own (see Technique LC [1634]). We will be referencing these results heavily in later sections, and will remind you then to come back for a second look.

Subsection READ Reading Questions

- 1. Why does Theorem G [852] have the title it does?
- 2. Why is Theorem RMRT [859] so surprising ?

3. Row-reduce the matrix A to reduced row-echelon form. Without any further computations, compute the dimensions of the four subspaces, (a) $\mathcal{N}(A)$, (b) $\mathcal{C}(A)$, (c) $\mathcal{R}(A)$ and (d) $\mathcal{L}(A)$.

$$A = \begin{bmatrix} 1 & -1 & 2 & 8 & 5\\ 1 & 1 & 1 & 4 & -1\\ 0 & 2 & -3 & -8 & -6\\ 2 & 0 & 1 & 8 & 4 \end{bmatrix}$$

Subsection EXC Exercises

C10 Example SVP4 [856] leaves several details for the reader to check. Verify these five claims.

Contributed by Robert Beezer

C40 Determine if the set $T = \{x^2 - x + 5, 4x^3 - x^2 + 5x, 3x + 2\}$ spans the vector space of polynomials with degree 4 or less, P_4 . (Compare the solution to this exercise with Solution LISS.C40 [772].) Contributed by Robert Beezer Solution [879]

M50 Mimic Definition DS [864] and construct a reasonable definition of $V = U_1 \oplus U_2 \oplus U_3 \oplus \ldots \oplus U_m$. Contributed by Robert Beezer

T05 Trivially, if U and V are two subspaces of W, then dim $(U) = \dim(V)$. Combine this fact, Theorem PSSD [857], and Theorem EDYES [858] all into one grand combined theorem. You might look to Theorem PIP [406] stylistic inspiration. (Notice this problem does not ask you to prove anything. It just asks you to roll up three theorems into one compact, logically equivalent statement.) Contributed by Robert Beezer

T10 Prove the following theorem, which could be viewed as a reformulation of parts (3) and (4) of Theorem G [852], or more appropriately as a corollary of Theorem G [852] (Technique LC [1634]).

Suppose V is a vector space and S is a subset of V such that the number of vectors in S equals the dimension of V. Then S is linearly independent if and only if S spans V. Contributed by Robert Beezer

T15 Suppose that A is an $m \times n$ matrix and let $\min(m, n)$ denote the minimum of m and n. Prove that $r(A) \leq \min(m, n)$. Contributed by Robert Beezer

T20 Suppose that A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{C}^m$. Prove that the linear system $\mathcal{LS}(A, \mathbf{b})$ is consistent if and only if $r(A) = r([A | \mathbf{b}])$. Contributed by Robert Beezer Solution [879]

T25 Suppose that V is a vector space with finite dimension. Let W be any subspace of V. Prove that W has finite dimension.

Contributed by Robert Beezer

T33 Part of Exercise B.T50 [804] is the half of the proof where we assume the matrix A is nonsingular and prove that a set is basis. In Solution B.T50 [811] we proved directly that the set was both linearly independent and a spanning set. Shorten this part of the proof by applying Theorem G [852]. Be careful, there is one subtlety. Contributed by Robert Beezer Solution [879]

T60 Suppose that W is a vector space with dimension 5, and U and V are subspaces of W, each of dimension 3. Prove that $U \cap V$ contains a non-zero vector. State a more general result.

Contributed by Joe Riegsecker Solution [880]

Subsection SOL Solutions

C40 Contributed by Robert Beezer Statement [876] The vector space P_4 has dimension 5 by Theorem DP [824]. Since T contains only 3 vectors, and 3 < 5, Theorem G [852] tells us that T does not span P_4 .

T20 Contributed by Robert Beezer Statement [877] (\Rightarrow) Suppose first that $\mathcal{LS}(A, \mathbf{b})$ is consistent. Then by Theorem CSCS [563], $\mathbf{b} \in \mathcal{C}(A)$. This means that $\mathcal{C}(A) = \mathcal{C}([A | \mathbf{b}])$ and so it follows that $r(A) = r([A | \mathbf{b}])$.

(\Leftarrow) Adding a column to a matrix will only increase the size of its column space, so in all cases, $\mathcal{C}(A) \subseteq \mathcal{C}([A | \mathbf{b}])$. However, if we assume that $r(A) = r([A | \mathbf{b}])$, then by Theorem EDYES [858] we conclude that $\mathcal{C}(A) = \mathcal{C}([A | \mathbf{b}])$. Then $\mathbf{b} \in \mathcal{C}([A | \mathbf{b}]) = \mathcal{C}(A)$ so by Theorem CSCS [563], $\mathcal{LS}(A, \mathbf{b})$ is consistent.

T33 Contributed by Robert Beezer Statement [877]

By Theorem DCM [824] we know that \mathbb{C}^n has dimension n. So by Theorem G [852] we need only establish that the set C is linearly independent or a spanning set. However, the hypotheses also require that C be of size n. We assumed that $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$ had size n, but there is no guarantee that $C = \{A\mathbf{x}_1, A\mathbf{x}_2, A\mathbf{x}_3, \ldots, A\mathbf{x}_n\}$ will have size n. There could be some "collapsing" or "collisions."

Suppose we establish that C is linearly independent. Then C must have n distinct

elements or else we could fashion a nontrivial relation of linear dependence involving duplicate elements.

If we instead to choose to prove that C is a spanning set, then we could establish the uniqueness of the elements of C quite easily. Suppose that $A\mathbf{x}_i = A\mathbf{x}_j$. Then

$$A(\mathbf{x}_i - \mathbf{x}_j) = A\mathbf{x}_i - A\mathbf{x}_j = \mathbf{0}$$

Since A is nonsingular, we conclude that $\mathbf{x}_i - \mathbf{x}_j = \mathbf{0}$, or $\mathbf{x}_i = \mathbf{x}_j$, contrary to our description of B.

T60 Contributed by Robert Beezer Statement [878]

Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be bases for U and V (respectively). Then, the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent, since Theorem G [852] says we cannot have 6 linearly independent vectors in a vector space of dimension 5. So we can assert that there is a non-trivial relation of linear dependence,

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 = \mathbf{0}$$

where a_1 , a_2 , a_3 and b_1 , b_2 , b_3 are not all zero.

We can rearrange this equation as

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 = -b_1\mathbf{v}_1 - b_2\mathbf{v}_2 - b_3\mathbf{v}_3$$

This is an equality of two vectors, so we can give this common vector a name, say \mathbf{w} ,

$$\mathbf{w} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 = -b_1\mathbf{v}_1 - b_2\mathbf{v}_2 - b_3\mathbf{v}_3$$

This is the desired non-zero vector, as we will now show.

First, since $\mathbf{w} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3$, we can see that $\mathbf{w} \in U$. Similarly, $\mathbf{w} = -b_1\mathbf{v}_1 - b_2\mathbf{v}_2 - b_3\mathbf{v}_3$, so $\mathbf{w} \in V$. This establishes that $\mathbf{w} \in U \cap V$ (Definition SI [1608]).

Is $\mathbf{w} \neq \mathbf{0}$? Suppose not, in other words, suppose $\mathbf{w} = \mathbf{0}$. Then

$$\mathbf{0} = \mathbf{w} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3$$

Because $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a basis for U, it is a linearly independent set and the relation of linear dependence above means we must conclude that $a_1 = a_2 = a_3 = 0$. By a similar process, we would conclude that $b_1 = b_2 = b_3 = 0$. But this is a contradiction since $a_1, a_2, a_3, b_1, b_2, b_3$ were chosen so that some were nonzero. So $\mathbf{w} \neq \mathbf{0}$.

How does this generalize? All we really needed was the original relation of linear dependence that resulted because we had "too many" vectors in W. A more general statement would be: Suppose that W is a vector space with dimension n, U is a subspace of dimension p and V is a subspace of dimension q. If p + q > n, then $U \cap V$ contains a non-zero vector.

Chapter D Determinants

The determinant is a function that takes a square matrix as an input and produces a scalar as an output. So unlike a vector space, it is not an algebraic structure. However, it has many beneficial properties for studying vector spaces, matrices and systems of equations, so it is hard to ignore (though some have tried). While the properties of a determinant can be very useful, they are also complicated to prove.

Section DM Determinant of a Matrix

First, a slight detour, as we introduce elementary matrices, which will bring us back to the beginning of the course and our old friend, row operations.

Subsection EM Elementary Matrices

Elementary matrices are very simple, as you might have suspected from their name. Their purpose is to effect row operations (Definition RO [59]) on a matrix through matrix multiplication (Definition MM [468]). Their definitions look much more complicated than they really are, so be sure to skip over them on your first reading and head right for the explanation that follows and the first example.

Definition ELEM Elementary Matrices 1. For $i \neq j$, $E_{i,j}$ is the square matrix of size n with

$$[E_{i,j}]_{k\ell} = \begin{cases} 0 & k \neq i, k \neq j, \ell \neq k \\ 1 & k \neq i, k \neq j, \ell = k \\ 0 & k = i, \ell \neq j \\ 1 & k = i, \ell = j \\ 0 & k = j, \ell \neq i \\ 1 & k = j, \ell = i \end{cases}$$

2. For $\alpha \neq 0$, $E_i(\alpha)$ is the square matrix of size n with

$$[E_i(\alpha)]_{k\ell} = \begin{cases} 0 & k \neq i, \ell \neq k \\ 1 & k \neq i, \ell = k \\ \alpha & k = i, \ell = i \end{cases}$$

3. For $i \neq j$, $E_{i,j}(\alpha)$ is the square matrix of size n with

$$\left[E_{i,j}(\alpha)\right]_{k\ell} = \begin{cases} 0 & k \neq j, \ell \neq k \\ 1 & k \neq j, \ell = k \\ 0 & k = j, \ell \neq i, \ell \neq j \\ 1 & k = j, \ell = j \\ \alpha & k = j, \ell = i \end{cases}$$

(This definition contains Notation ELEM.)

 \triangle

Again, these matrices are not as complicated as their definitions suggest, since they are just small perturbations of the $n \times n$ identity matrix (Definition IM [175]). $E_{i,j}$ is the identity matrix with rows (or columns) i and j trading places, $E_i(\alpha)$ is the identity matrix where the diagonal entry in row i and column i has been replaced by α , and $E_{i,j}(\alpha)$ is the identity matrix where the entry in row j and column i has been replaced by α . (Yes, those subscripts look backwards in the description of $E_{i,j}(\alpha)$). Notice that our notation makes no reference to the size of the elementary matrix, since this will always be apparent from the context, or unimportant.

The *raison d'être* for elementary matrices is to "do" row operations on matrices with matrix multiplication. So here is an example where we will both see some elementary matrices and see how they accomplish row operations when used with matrix multiplication.

Example EMRO

Elementary matrices and row operations

We will perform a sequence of row operations (Definition RO [59]) on the 3×4 matrix A, while also multiplying the matrix on the left by the appropriate 3×3 elementary matrix.

$$A = \begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & 3 & 2 & 4 \\ 5 & 0 & 3 & 1 \end{bmatrix}$$

$$R_{1} \leftrightarrow R_{3}: \begin{bmatrix} 5 & 0 & 3 & 1 \\ 1 & 3 & 2 & 4 \\ 2 & 1 & 3 & 1 \end{bmatrix} \qquad E_{1,3}: \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & 3 & 2 & 4 \\ 5 & 0 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 3 & 1 \\ 1 & 3 & 2 & 4 \\ 2 & 1 & 3 & 1 \end{bmatrix}$$
$$2R_{2}: \begin{bmatrix} 5 & 0 & 3 & 1 \\ 2 & 6 & 4 & 8 \\ 2 & 1 & 3 & 1 \end{bmatrix} \qquad E_{2}(2): \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 3 & 1 \\ 1 & 3 & 2 & 4 \\ 2 & 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 3 & 1 \\ 2 & 6 & 4 & 8 \\ 2 & 1 & 3 & 1 \end{bmatrix}$$
$$2R_{3} + R_{1}: \begin{bmatrix} 9 & 2 & 9 & 3 \\ 2 & 6 & 4 & 8 \\ 2 & 1 & 3 & 1 \end{bmatrix} \qquad E_{3,1}(2): \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 3 & 1 \\ 2 & 6 & 4 & 8 \\ 2 & 1 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 9 & 2 & 9 & 3 \\ 2 & 6 & 4 & 8 \\ 2 & 1 & 3 & 1 \end{bmatrix}$$

The next three theorems establish that each elementary matrix effects a row operation via matrix multiplication.

Theorem EMDRO Elementary Matrices Do Row Operations

Suppose that A is an $m \times n$ matrix, and B is a matrix of the same size that is obtained from A by a single row operation (Definition RO [59]). Then there is an elementary matrix of size m that will convert A to B via matrix multiplication on the left. More precisely,

- 1. If the row operation swaps rows i and j, then $B = E_{i,j}A$.
- 2. If the row operation multiplies row *i* by α , then $B = E_i(\alpha) A$.

 \boxtimes

3. If the row operation multiplies row *i* by α and adds the result to row *j*, then $B = E_{i,j}(\alpha) A$.

Proof In each of the three conclusions, performing the row operation on A will create the matrix B where only one or two rows will have changed. So we will establish the equality of the matrix entries row by row, first for the unchanged rows, then for the changed rows, showing in each case that the result of the matrix product is the same as the result of the row operation. Here we go.

Row k of the product $E_{i,j}A$, where $k \neq i, k \neq j$, is unchanged from A,

$$\begin{split} [E_{i,j}A]_{k\ell} &= \sum_{p=1}^{n} [E_{i,j}]_{kp} [A]_{p\ell} & \text{Theorem EMP [471]} \\ &= [E_{i,j}]_{kk} [A]_{k\ell} + \sum_{\substack{p=1\\p \neq k}}^{n} [E_{i,j}]_{kp} [A]_{p\ell} & \text{Property CACN [1597]} \\ &= 1 [A]_{k\ell} + \sum_{\substack{p=1\\p \neq k}}^{n} 0 [A]_{p\ell} & \text{Definition ELEM [883]} \\ &= [A]_{k\ell} \end{split}$$

Row *i* of the product $E_{i,j}A$ is row *j* of *A*,

$$[E_{i,j}A]_{i\ell} = \sum_{p=1}^{n} [E_{i,j}]_{ip} [A]_{p\ell}$$

= $[E_{i,j}]_{ij} [A]_{j\ell} + \sum_{\substack{p=1\\p\neq j}}^{n} [E_{i,j}]_{ip} [A]_{p\ell}$
= $1 [A]_{j\ell} + \sum_{\substack{p=1\\p\neq j}}^{n} 0 [A]_{p\ell}$
= $[A]_{j\ell}$

Theorem EMP [471]

Property CACN [1597]

Definition ELEM [883]

Row j of the product $E_{i,j}A$ is row i of A,

$$[E_{i,j}A]_{j\ell} = \sum_{p=1}^{n} [E_{i,j}]_{jp} [A]_{p\ell}$$

= $[E_{i,j}]_{ji} [A]_{i\ell} + \sum_{\substack{p=1\\p\neq i}}^{n} [E_{i,j}]_{jp} [A]_{p\ell}$
= $1 [A]_{i\ell} + \sum_{\substack{p=1\\p\neq i}}^{n} 0 [A]_{p\ell}$

Theorem EMP [471]

Property CACN [1597]

Definition ELEM [883]

$$= [A]_{i\ell}$$

L

So the matrix product $E_{i,j}A$ is the same as the row operation that swaps rows *i* and *j*. Row *k* of the product $E_i(\alpha)A$, where $k \neq i$, is unchanged from *A*,

$$E_{i}(\alpha) A]_{k\ell} = \sum_{p=1}^{n} [E_{i}(\alpha)]_{kp} [A]_{p\ell} \qquad \text{Theorem EMP [471]}$$

$$= [E_{i}(\alpha)]_{kk} [A]_{k\ell} + \sum_{\substack{p=1\\p\neq k}}^{n} [E_{i}(\alpha)]_{kp} [A]_{p\ell} \qquad \text{Property CACN [1597]}$$

$$= 1 [A]_{k\ell} + \sum_{\substack{p=1\\p\neq k}}^{n} 0 [A]_{p\ell} \qquad \text{Definition ELEM [883]}$$

$$= [A]_{k\ell}$$

Row *i* of the product $E_i(\alpha) A$ is α times row *i* of *A*,

$$[E_{i}(\alpha) A]_{i\ell} = \sum_{p=1}^{n} [E_{i}(\alpha)]_{ip} [A]_{p\ell}$$
 Theorem EMP [471]
$$= [E_{i}(\alpha)]_{ii} [A]_{i\ell} + \sum_{\substack{p=1\\p\neq i}}^{n} [E_{i}(\alpha)]_{ip} [A]_{p\ell}$$
 Property CACN [1597]

$$= \alpha \left[A \right]_{i\ell} + \sum_{\substack{p=1\\p \neq i}}^{n} 0 \left[A \right]_{p\ell}$$
$$= \alpha \left[A \right]_{i\ell}$$

Definition ELEM [883]

So the matrix product $E_i(\alpha) A$ is the same as the row operation that swaps multiplies row i by α .

Row k of the product $E_{i,j}(\alpha) A$, where $k \neq j$, is unchanged from A,

$$\begin{split} \left[E_{i,j}\left(\alpha\right)A\right]_{k\ell} &= \sum_{p=1}^{n} \left[E_{i,j}\left(\alpha\right)\right]_{kp} \left[A\right]_{p\ell} & \text{Theorem EMP [471]} \\ &= \left[E_{i,j}\left(\alpha\right)\right]_{kk} \left[A\right]_{k\ell} + \sum_{\substack{p=1\\p \neq k}}^{n} \left[E_{i,j}\left(\alpha\right)\right]_{kp} \left[A\right]_{p\ell} & \text{Property CACN [1597]} \\ &= 1 \left[A\right]_{k\ell} + \sum_{\substack{p=1\\p \neq k}}^{n} 0 \left[A\right]_{p\ell} & \text{Definition ELEM [883]} \\ &= \left[A\right]_{k\ell} \end{split}$$

Row j of the product $E_{i,j}(\alpha) A$, is α times row i of A and then added to row j of A,

$$[E_{i,j}(\alpha)A]_{j\ell} = \sum_{p=1}^{n} [E_{i,j}(\alpha)]_{jp} [A]_{p\ell}$$
 Theorem EMP [471]

$$= [E_{i,j} (\alpha)]_{jj} [A]_{j\ell} + \sum_{\substack{p=1\\p\neq j,i}}^{n} [E_{i,j} (\alpha)]_{jp} [A]_{p\ell} \qquad \text{Property CACN [1597]}$$
$$= 1 [A]_{j\ell} + \alpha [A]_{i\ell} + \sum_{\substack{p=1\\p\neq j,i}}^{n} 0 [A]_{p\ell} \qquad \text{Definition ELEM [883]}$$
$$= [A]_{j\ell} + \alpha [A]_{i\ell}$$

So the matrix product $E_{i,j}(\alpha) A$ is the same as the row operation that multiplies row *i* by α and adds the result to row *j*.

Later in this section we will need two facts about elementary matrices.

Theorem EMN Elementary Matrices are Nonsingular

If E is an elementary matrix, then E is nonsingular.

Proof We show that we can row-reduce each elementary matrix to the identity matrix. Given an elementary matrix of the form $E_{i,j}$, perform the row operation that swaps row j with row i. Given an elementary matrix of the form $E_i(\alpha)$, with $\alpha \neq 0$, perform the row operation that multiplies row i by $1/\alpha$. Given an elementary matrix of the form $E_{i,j}(\alpha)$, with $\alpha \neq 0$, perform the row operation that multiplies row i by $1/\alpha$. Given an elementary matrix of the form $E_{i,j}(\alpha)$, with $\alpha \neq 0$, perform the row operation that multiplies row i by $-\alpha$ and adds it to row j. In each case, the result of the single row operation is the identity matrix. So each elementary matrix

 \square

is row-equivalent to the identity matrix, and by Theorem NMRRI [176] is nonsingular.

Notice that we have now made use of the nonzero restriction on α in the definition of $E_i(\alpha)$. One more key property of elementary matrices.

Theorem NMPEM

Nonsingular Matrices are Products of Elementary Matrices

Suppose that A is a nonsingular matrix. Then there exists elementary matrices $E_1, E_2, E_3, \ldots, E_t$ so that $A = E_1 E_2 E_3 \ldots E_t$.

Proof Since A is nonsingular, it is row-equivalent to the identity matrix by Theorem NMRRI [176], so there is a sequence of t row operations that converts I to A. For each of these row operations, form the associated elementary matrix from Theorem EMDRO [886] and denote these matrices by $E_1, E_2, E_3, \ldots, E_t$. Applying the first row operation to I yields the matrix E_1I . The second row operation yields $E_2(E_1I)$, and the third row operation creates $E_3E_2E_1I$. The result of the full sequence of t row operations will yield A, so

$$A = E_t \dots E_3 E_2 E_1 I = E_t \dots E_3 E_2 E_1$$

Other than the cosmetic matter of re-indexing these elementary matrices in the opposite order, this is the desired result.

Subsection DD Definition of the Determinant

We'll now turn to the definition of a determinant and do some sample computations. The definition of the determinant function is **recursive**, that is, the determinant of a large matrix is defined in terms of the determinant of smaller matrices. To this end, we will make a few definitions.

Definition SM SubMatrix

Suppose that A is an $m \times n$ matrix. Then the **submatrix** A(i|j) is the $(m-1) \times (n-1)$ matrix obtained from A by removing row i and column j.

(This definition contains Notation SM.)

 \triangle

Example SS Some submatrices For the matrix

$$A = \begin{bmatrix} 1 & -2 & 3 & 9 \\ 4 & -2 & 0 & 1 \\ 3 & 5 & 2 & 1 \end{bmatrix}$$

we have the submatrices

$$A(2|3) = \begin{bmatrix} 1 & -2 & 9 \\ 3 & 5 & 1 \end{bmatrix} \qquad A(3|1) = \begin{bmatrix} -2 & 3 & 9 \\ -2 & 0 & 1 \end{bmatrix}$$

Definition DM Determinant of a Matrix

Suppose A is a square matrix. Then its **determinant**, det (A) = |A|, is an element of \mathbb{C} defined recursively by:

If A is a 1×1 matrix, then det $(A) = [A]_{11}$.

If A is a matrix of size n with $n \ge 2$, then

$$\det (A) = [A]_{11} \det (A(1|1)) - [A]_{12} \det (A(1|2)) + [A]_{13} \det (A(1|3)) - [A]_{14} \det (A(1|4)) + \dots + (-1)^{n+1} [A]_{1n} \det (A(1|n))$$

(This definition contains Notation DM.)

So to compute the determinant of a 5×5 matrix we must build 5 submatrices, each of size 4. To compute the determinants of each the 4×4 matrices we need to create 4 submatrices each, these now of size 3 and so on. To compute the determinant of a 10×10 matrix would require computing the determinant of $10! = 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 = 3,628,800$

 \triangle

 \boxtimes

 1×1 matrices. Fortunately there are better ways. However this does suggest an excellent computer programming exercise to write a recursive procedure to compute a determinant.

Let's compute the determinant of a reasonable sized matrix by hand.

Example D33M Determinant of a 3×3 matrix Suppose that we have the 3×3 matrix

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 4 & 1 & 6 \\ -3 & -1 & 2 \end{bmatrix}$$

Then

$$det (A) = |A| = \begin{vmatrix} 3 & 2 & -1 \\ 4 & 1 & 6 \\ -3 & -1 & 2 \end{vmatrix}$$
$$= 3 \begin{vmatrix} 1 & 6 \\ -1 & 2 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ -3 & 2 \end{vmatrix} + (-1) \begin{vmatrix} 4 & 1 \\ -3 & -1 \end{vmatrix}$$
$$= 3 (1 |2| - 6 |-1|) - 2 (4 |2| - 6 |-3|) - (4 |-1| - 1 |-3|)$$
$$= 3 (1(2) - 6(-1)) - 2 (4(2) - 6(-3)) - (4(-1) - 1(-3))$$
$$= 24 - 52 + 1$$
$$= -27$$

 \boxtimes

 \square

In practice it is a bit silly to decompose a 2×2 matrix down into a couple of 1×1 matrices and then compute the exceedingly easy determinant of these puny matrices. So here is a simple theorem.

Theorem DMST Determinant of Matrices of Size Two Suppose that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then det (A) = ad - bc.

Proof Applying Definition DM [894],

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a \left| d \right| - b \left| c \right| = ad - bc$$

Do you recall seeing the expression ad - bc before? (Hint: Theorem TTMI [509])

Subsection CD Computing Determinants

There are a variety of ways to compute the determinant. We will establish first that we can choose to mimic our definition of the determinant, but by using matrix entries and

submatrices based on a row other than the first one.

Theorem DER Determinant Expansion about Rows

Suppose that A is a square matrix of size n. Then

$$\det (A) = (-1)^{i+1} [A]_{i1} \det (A(i|1)) + (-1)^{i+2} [A]_{i2} \det (A(i|2)) + (-1)^{i+3} [A]_{i3} \det (A(i|3)) + \dots + (-1)^{i+n} [A]_{in} \det (A(i|n)) \qquad 1 \le i \le n$$

which is known as **expansion** about row i.

Proof First, the statement of the theorem coincides with Definition DM [894] when i = 1, so throughout, we need only consider i > 1.

Given the recursive definition of the determinant, it should be no surprise that we will use induction for this proof (Technique I [1630]). When n = 1, there is nothing to prove since there is but one row. When n = 2, we just examine expansion about the second row,

$$\begin{aligned} (-1)^{2+1} [A]_{21} \det (A(2|1)) + (-1)^{2+2} [A]_{22} \det (A(2|2)) \\ &= -[A]_{21} [A]_{12} + [A]_{22} [A]_{11} \\ &= [A]_{11} [A]_{22} - [A]_{12} [A]_{21} \\ &= \det (A) \end{aligned}$$
 Definition DM [894]
Theorem DMST [896]

So the theorem is true for matrices of size n = 1 and n = 2. Now assume the result is true for all matrices of size n - 1 as we derive an expression for expansion about row *i* for a matrix of

size n. We will abuse our notation for a submatrix slightly, so $A(i_1, i_2|j_1, j_2)$ will denote the matrix formed by removing rows i_1 and i_2 , along with removing columns j_1 and j_2 . Also, as we take a determinant of a submatrix, we will need to "jump up" the index of summation partway through as we "skip over" a missing column. To do this smoothly we will set

$$\epsilon_{\ell j} = \begin{cases} 0 & \ell < j \\ 1 & \ell > j \end{cases}$$

Now,

$$\det (A) = \sum_{j=1}^{n} (-1)^{1+j} [A]_{1j} \det (A(1|j))$$
Definition DM [894]
$$= \sum_{j=1}^{n} (-1)^{1+j} [A]_{1j} \sum_{\substack{1 \le \ell \le n \\ \ell \ne j}} (-1)^{i-1+\ell-\epsilon_{\ell j}} [A]_{i\ell} \det (A(1,i|j,\ell))$$
Induction Hypothesis
$$= \sum_{j=1}^{n} \sum_{\substack{1 \le \ell \le n \\ \ell \ne j}} (-1)^{j+i+\ell-\epsilon_{\ell j}} [A]_{1j} [A]_{i\ell} \det (A(1,i|j,\ell))$$
Property DCN [1597]
$$= \sum_{\ell=1}^{n} \sum_{\substack{1 \le j \le n \\ j \ne \ell}} (-1)^{j+i+\ell-\epsilon_{\ell j}} [A]_{1j} [A]_{i\ell} \det (A(1,i|j,\ell))$$
Property CACN [1597]

$$= \sum_{\ell=1}^{n} (-1)^{i+\ell} [A]_{i\ell} \sum_{\substack{1 \le j \le n \\ j \ne \ell}} (-1)^{j-\epsilon_{\ell j}} [A]_{1j} \det (A(1,i|j,\ell))$$
Property DCN [1597]
$$= \sum_{\ell=1}^{n} (-1)^{i+\ell} [A]_{i\ell} \sum_{\substack{1 \le j \le n \\ j \ne \ell}} (-1)^{\epsilon_{\ell j}+j} [A]_{1j} \det (A(i,1|\ell,j))$$
2\epsilon_{\ell j} is even
$$= \sum_{\ell=1}^{n} (-1)^{i+\ell} [A]_{i\ell} \det (A(i|\ell))$$
Definition DM [894]

We can also obtain a formula that computes a determinant by expansion about a column, but this will be simpler if we first prove a result about the interplay of determinants and transposes. Notice how the following proof makes use of the ability to compute a determinant by expanding about *any* row.

Theorem DT Determinant of the Transpose

Suppose that A is a square matrix. Then $\det(A^t) = \det(A)$.

Proof With our definition of the determinant (Definition DM [894]) and theorems like Theorem DER [897], using induction (Technique I [1630]) is a natural approach to proving properties of determinants. And so it is here. Let n be the size of the matrix A, and we will use induction on n.

 \square

For n = 1, the transpose of a matrix is identical to the original matrix, so vacuously, the determinants are equal.

Now assume the result is true for matrices of size n-1. Then,

$$\det (A^{t}) = \frac{1}{n} \sum_{i=1}^{n} \det (A^{t})$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} [A^{t}]_{ij} \det (A^{t}(i|j)) \qquad \text{Theorem DER [897]}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} [A]_{ji} \det (A^{t}(i|j)) \qquad \text{Definition TM [433]}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} [A]_{ji} \det ((A(j|i))^{t}) \qquad \text{Definition TM [433]}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (-1)^{i+j} [A]_{ji} \det (A(j|i)) \qquad \text{Induction Hypothesis}$$

$$= \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} (-1)^{j+i} [A]_{ji} \det (A(j|i)) \qquad \text{Property CACN [1597]}$$

$$= \frac{1}{n} \sum_{j=1}^{n} \det (A) \qquad \text{Theorem DER [897]}$$

$$= \det(A)$$

Now we can easily get the result that a determinant can be computed by expansion about any column as well.

Theorem DEC

Determinant Expansion about Columns

Suppose that A is a square matrix of size n. Then

$$det (A) = (-1)^{1+j} [A]_{1j} det (A (1|j)) + (-1)^{2+j} [A]_{2j} det (A (2|j)) + (-1)^{3+j} [A]_{3j} det (A (3|j)) + \dots + (-1)^{n+j} [A]_{nj} det (A (n|j)) \qquad 1 \le j \le n$$

which is known as **expansion** about column j.

Proof

$$\det (A) = \det (A^{t}) \qquad \text{Theorem DT [899]}$$

$$= \sum_{i=1}^{n} (-1)^{j+i} [A^{t}]_{ji} \det (A^{t} (j|i)) \qquad \text{Theorem DER [897]}$$

$$= \sum_{i=1}^{n} (-1)^{j+i} [A^{t}]_{ji} \det ((A (i|j))^{t}) \qquad \text{Definition TM [433]}$$

$$= \sum_{i=1}^{n} (-1)^{j+i} [A^{t}]_{ji} \det (A (i|j)) \qquad \text{Theorem DT [899]}$$

$$=\sum_{i=1}^{n} (-1)^{i+j} [A]_{ij} \det (A(i|j))$$
 Definition TM [433]

That the determinant of an $n \times n$ matrix can be computed in 2n different (albeit similar) ways is nothing short of remarkable. For the doubters among us, we will do an example, computing a 4×4 matrix in two different ways.

Example TCSD Two computations, same determinant Let

$$A = \begin{bmatrix} -2 & 3 & 0 & 1\\ 9 & -2 & 0 & 1\\ 1 & 3 & -2 & -1\\ 4 & 1 & 2 & 6 \end{bmatrix}$$

Then expanding about the fourth row (Theorem DER [897] with i = 4) yields,

$$|A| = (4)(-1)^{4+1} \begin{vmatrix} 3 & 0 & 1 \\ -2 & 0 & 1 \\ 3 & -2 & -1 \end{vmatrix} + (1)(-1)^{4+2} \begin{vmatrix} -2 & 0 & 1 \\ 9 & 0 & 1 \\ 1 & -2 & -1 \end{vmatrix} + (2)(-1)^{4+3} \begin{vmatrix} -2 & 3 & 1 \\ 9 & -2 & 1 \\ 1 & 3 & -1 \end{vmatrix} + (6)(-1)^{4+4} \begin{vmatrix} -2 & 3 & 0 \\ 9 & -2 & 0 \\ 1 & 3 & -2 \end{vmatrix}$$

$$= (-4)(10) + (1)(-22) + (-2)(61) + 6(46) = 92$$

while expanding about column 3 (Theorem DEC [901] with j = 3) gives

$$|A| = (0)(-1)^{1+3} \begin{vmatrix} 9 & -2 & 1 \\ 1 & 3 & -1 \\ 4 & 1 & 6 \end{vmatrix} + (0)(-1)^{2+3} \begin{vmatrix} -2 & 3 & 1 \\ 1 & 3 & -1 \\ 4 & 1 & 6 \end{vmatrix} + (-2)(-1)^{3+3} \begin{vmatrix} -2 & 3 & 1 \\ 9 & -2 & 1 \\ 4 & 1 & 6 \end{vmatrix} + (2)(-1)^{4+3} \begin{vmatrix} -2 & 3 & 1 \\ 9 & -2 & 1 \\ 1 & 3 & -1 \end{vmatrix} = 0 + 0 + (-2)(-107) + (-2)(61) = 92$$

Notice how much easier the second computation was. By choosing to expand about the third column, we have two entries that are zero, so two 3×3 determinants need not be computed at all!

When a matrix has all zeros above (or below) the diagonal, exploiting the zeros by expanding about the proper row or column makes computing a determinant insanely easy.

Example DUTM Determinant of an upper triangular matrix Suppose that

$$T = \begin{bmatrix} 2 & 3 & -1 & 3 & 3 \\ 0 & -1 & 5 & 2 & -1 \\ 0 & 0 & 3 & 9 & 2 \\ 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

We will compute the determinant of this 5×5 matrix by consistently expanding about the first column for each submatrix that arises and does not have a zero entry multiplying it.

$$\det (T) = \begin{vmatrix} 2 & 3 & -1 & 3 & 3 \\ 0 & -1 & 5 & 2 & -1 \\ 0 & 0 & 3 & 9 & 2 \\ 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 5 \end{vmatrix}$$
$$= 2(-1)^{1+1} \begin{vmatrix} -1 & 5 & 2 & -1 \\ 0 & 3 & 9 & 2 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{vmatrix}$$
$$= 2(-1)(-1)^{1+1} \begin{vmatrix} 3 & 9 & 2 \\ 0 & -1 & 3 \\ 0 & 0 & 5 \end{vmatrix}$$
$$= 2(-1)(3)(-1)^{1+1} \begin{vmatrix} -1 & 3 \\ 0 & 5 \end{vmatrix}$$

$$= 2(-1)(3)(-1)(-1)^{1+1} |5|$$

= 2(-1)(3)(-1)(5) = 30

 \boxtimes

If you consult other texts in your study of determinants, you may run into the terms "minor" and "cofactor," especially in a discussion centered on expansion about rows and columns. We've chosen not to make these definitions formally since we've been able to get along without them. However, informally, a **minor** is a determinant of a submatrix, specifically det (A(i|j)) and is usually referenced as the minor of $[A]_{ij}$. A **cofactor** is a signed minor, specifically the cofactor of $[A]_{ij}$ is $(-1)^{i+j} \det (A(i|j))$.

Subsection READ Reading Questions

- 1. Construct the elementary matrix that will effect the row operation $-6R_2 + R_3$ on a 4×7 matrix.
- 2. Compute the determinant of the matrix

$$\begin{bmatrix} 2 & 3 & -1 \\ 3 & 8 & 2 \\ 4 & -1 & -3 \end{bmatrix}$$

3. Compute the determinant of the matrix

$$\begin{bmatrix} 3 & 9 & -2 & 4 & 2 \\ 0 & 1 & 4 & -2 & 7 \\ 0 & 0 & -2 & 5 & 2 \\ 0 & 0 & 0 & -1 & 6 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Subsection EXC Exercises

C21 Doing the computations by hand, find the determinant of the matrix below.

$$\begin{bmatrix} 1 & 3 \\ 6 & 2 \end{bmatrix}$$

Contributed by Chris Black Solution [912]

C22 Doing the computations by hand, find the determinant of the matrix below.

$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

Contributed by Chris Black Solution [912]

C23 Doing the computations by hand, find the determinant of the matrix below.

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix}$$

Contributed by Chris Black Solution [912]

C24 Doing the computations by hand, find the determinant of the matrix below.

$$\begin{bmatrix} -2 & 3 & -2 \\ -4 & -2 & 1 \\ 2 & 4 & 2 \end{bmatrix}$$

Contributed by Robert Beezer Solution [913]

C25 Doing the computations by hand, find the determinant of the matrix below.

$$\begin{bmatrix} 3 & -1 & 4 \\ 2 & 5 & 1 \\ 2 & 0 & 6 \end{bmatrix}$$

Contributed by Robert Beezer Solution [913]

C26 Doing the computations by hand, find the determinant of the matrix A.

$$A = \begin{bmatrix} 2 & 0 & 3 & 2 \\ 5 & 1 & 2 & 4 \\ 3 & 0 & 1 & 2 \\ 5 & 3 & 2 & 1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [914]

C27 Doing the computations by hand, find the determinant of the matrix A.

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & 2 & -1 & 1 \\ 2 & 1 & 3 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

Contributed by Chris Black Solution [914]

C28 Doing the computations by hand, find the determinant of the matrix A.

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ 2 & 5 & 3 & 0 \\ 1 & -1 & 0 & 1 \end{bmatrix}$$

Contributed by Chris Black Solution [915]

C29 Doing the computations by hand, find the determinant of the matrix A.

$$A = \begin{bmatrix} 2 & 3 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}$$

Contributed by Chris Black Solution [915]

C30 Doing the computations by hand, find the determinant of the matrix A.

$$A = \begin{bmatrix} 2 & 1 & 1 & 0 & 1 \\ 2 & 1 & 2 & -1 & 1 \\ 0 & 0 & 1 & 2 & 0 \\ 1 & 0 & 3 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 \end{bmatrix}$$

Contributed by Chris Black Solution [916]

M10 Find a value of k so that the matrix $A = \begin{bmatrix} 2 & 4 \\ 3 & k \end{bmatrix}$ has det(A) = 0, or explain why it is not possible. Contributed by Chris Black Solution [917] **M11** Find a value of k so that the matrix $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 1 \\ 2 & 3 & k \end{bmatrix}$ has $\det(A) = 0$, or explain why

it is not possible.

Contributed by Chris Black Solution [917]

M15 Given the matrix $B = \begin{bmatrix} 2-x & 1 \\ 4 & 2-x \end{bmatrix}$, find all values of x that are solutions of $\det(B) = 0$. Contributed by Chris Black Solution [917]

M16 Given the matrix
$$B = \begin{bmatrix} 4-x & -4 & -4 \\ 2 & -2-x & -4 \\ 3 & -3 & -4-x \end{bmatrix}$$
, find all values of x that are solu-

tions of det(B) = 0. Contributed by Chris Black Solution [918]

Subsection SOL Solutions

C21 Contributed by Chris Black Statement [907] Using the formula in Theorem DMST [896] we have

$$\begin{vmatrix} 1 & 3 \\ 6 & 2 \end{vmatrix} = 1 \cdot 2 - 6 \cdot 3 = 2 - 18 = -16$$

C22 Contributed by Chris Black Statement [907] Using the formula in Theorem DMST [896] we have

$$\begin{vmatrix} 1 & 3 \\ 2 & 6 \end{vmatrix} = 1 \cdot 6 - 2 \cdot 3 = 6 - 6 = 0$$

C23 Contributed by Chris Black Statement [907]

We can compute the determinant by expanding about any row or column; the most efficient ones to choose are either the second column or the third row. In any case, the determinant will be -4.

C24 Contributed by Robert Beezer Statement [908]

We'll expand about the first row since there are no zeros to exploit,

$$\begin{vmatrix} -2 & 3 & -2 \\ -4 & -2 & 1 \\ 2 & 4 & 2 \end{vmatrix} = (-2)\begin{vmatrix} -2 & 1 \\ 4 & 2 \end{vmatrix} + (-1)(3)\begin{vmatrix} -4 & 1 \\ 2 & 2 \end{vmatrix} + (-2)\begin{vmatrix} -4 & -2 \\ 2 & 4 \end{vmatrix}$$
$$= (-2)((-2)(2) - 1(4)) + (-3)((-4)(2) - 1(2)) + (-2)((-4)(4) - (-2)(2))$$
$$= (-2)(-8) + (-3)(-10) + (-2)(-12) = 70$$

C25 Contributed by Robert Beezer Statement [908]

We can expand about any row or column, so the zero entry in the middle of the last row is attractive. Let's expand about column 2. By Theorem DER [897] and Theorem DEC [901] you will get the same result by expanding about a different row or column. We will use Theorem DMST [896] twice.

$$\begin{vmatrix} 3 & -1 & 4 \\ 2 & 5 & 1 \\ 2 & 0 & 6 \end{vmatrix} = (-1)(-1)^{1+2} \begin{vmatrix} 2 & 1 \\ 2 & 6 \end{vmatrix} + (5)(-1)^{2+2} \begin{vmatrix} 3 & 4 \\ 2 & 6 \end{vmatrix} + (0)(-1)^{3+2} \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix}$$
$$= (1)(10) + (5)(10) + 0 = 60$$

C26 Contributed by Robert Beezer Statement [908]

With two zeros in column 2, we choose to expand about that column (Theorem DEC [901]),

$$\det (A) = \begin{vmatrix} 2 & 0 & 3 & 2 \\ 5 & 1 & 2 & 4 \\ 3 & 0 & 1 & 2 \\ 5 & 3 & 2 & 1 \end{vmatrix}$$
$$= 0(-1) \begin{vmatrix} 5 & 2 & 4 \\ 3 & 1 & 2 \\ 5 & 2 & 1 \end{vmatrix} + 1(1) \begin{vmatrix} 2 & 3 & 2 \\ 3 & 1 & 2 \\ 5 & 2 & 1 \end{vmatrix} + 0(-1) \begin{vmatrix} 2 & 3 & 2 \\ 5 & 2 & 4 \\ 5 & 2 & 1 \end{vmatrix} + 3(1) \begin{vmatrix} 2 & 3 & 2 \\ 5 & 2 & 4 \\ 3 & 1 & 2 \end{vmatrix}$$
$$= (1) (2(1(1) - 2(2)) - 3(3(1) - 5(2)) + 2(3(2) - 5(1))) + (3) (2(2(2) - 4(1)) - 3(5(2) - 4(3)) + 2(5(1) - 3(2)))$$
$$= (-6 + 21 + 2) + (3)(0 + 6 - 2) = 29$$

C27 Contributed by Chris Black Statement [909] Expanding on the first row, we have

$$\begin{vmatrix} 1 & 0 & 1 & 1 \\ 2 & 2 & -1 & 1 \\ 2 & 1 & 3 & 0 \\ 1 & 1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 2 & -1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{vmatrix} - 0 + \begin{vmatrix} 2 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix} - \begin{vmatrix} 2 & 2 & -1 \\ 2 & 1 & 3 \\ 1 & 1 & 0 \end{vmatrix}$$
$$= 4 + (-1) - (-1) = 4$$

C28 Contributed by Chris Black Statement [909] Expanding along the first row, we have

$$\begin{vmatrix} 1 & 0 & 1 & 1 \\ 2 & -1 & -1 & 1 \\ 2 & 5 & 3 & 0 \\ 1 & -1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} -1 & -1 & 1 \\ 5 & 3 & 0 \\ -1 & 0 & 1 \end{vmatrix} - 0 + \begin{vmatrix} 2 & -1 & 1 \\ 2 & 5 & 0 \\ 1 & -1 & 1 \end{vmatrix} - \begin{vmatrix} 2 & -1 & -1 \\ 2 & 5 & 3 \\ 1 & -1 & 0 \end{vmatrix}$$
$$= 5 - 0 + 5 - 10 = 0.$$

C29 Contributed by Chris Black Statement [910] Expanding along the first column, we have

$$\begin{vmatrix} 2 & 3 & 0 & 2 & 1 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & 2 & 3 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 \end{vmatrix} + 0 + 0 + 0 + 0$$

Now, expanding along the first column again, we have

$$= 2 \left(\begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 0 & 1 & 2 \end{vmatrix} - 0 + \begin{vmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 0 & 1 & 2 \end{vmatrix} - 0 \right)$$

$$= 2([2+0+6-0-0-8] + [4+0+2-0-3-2])$$

= 2

C30 Contributed by Chris Black Statement [910] In order to exploit the zeros, let's expand along row 3. We then have

2	3	0	2	1		10	1	0	1		10	1	1	1
0	1	1	1	2	(1)6		T	0	1	$+(-1)^{7}\cdot 2$		T	T	
0	0	1	- -	_		2	1	-1	1		2	1	2	1
0	0	T	Ζ	0	$ = (-1)^{*}$	1	0	1	1	$+(-1)^{-1} \cdot 2$	1	0	3	1
1	0	3	1	1			1	1	1			1	1	1
2	1	1	2	1		2	T	2	1		2	T	T	1
-	т	т	4	т										

Notice that the second matrix here is singular since two rows are identical and thus it cannot row-reduce to an identity matrix. We now have

$$= \begin{vmatrix} 2 & 1 & 0 & 1 \\ 2 & 1 & -1 & 1 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 2 & 1 \end{vmatrix} + 0$$

and now we expand on the first row of the first matrix:

$$= 2 \begin{vmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{vmatrix} - \begin{vmatrix} 2 & -1 & 1 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{vmatrix} + 0 - \begin{vmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ 2 & 1 & 2 \end{vmatrix}$$

$$= 2(-3) - (-3) - (-3) = 0$$

M10 Contributed by Chris Black Statement [910] There is only one value of k that will make this matrix have a zero determinant.

$$\det\left(A\right) = \begin{vmatrix} 2 & 4 \\ 3 & k \end{vmatrix} = 2k - 12$$

so det (A) = 0 only when k = 6.

M11 Contributed by Chris Black Statement [911]

$$\det(A) = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 0 & 1 \\ 2 & 3 & k \end{vmatrix} = 7 - 4k$$

Thus, $\det(A) = 0$ only when $k = \frac{7}{4}$.

M15 Contributed by Chris Black Statement [911] Using the formula for the determinant of a 2×2 matrix given in Theorem DMST [896], we have

$$\det(B) = \begin{vmatrix} 2-x & 1\\ 4 & 2-x \end{vmatrix} = (2-x)(2-x) - 4 = x^2 - 4x = x(x-4)$$

and thus det (B) = 0 only when x = 0 or x = 4.

M16 Contributed by Chris Black Statement [911]

$$\det(B) = 8x - 2x^2 - x^3 = -x(x^2 + 2x - 8) = -x(x - 2)(x + 4)$$

And thus, det(B) = 0 when x = 0, x = 2, or x = -4.

Section PDM Properties of Determinants of Matrices

We have seen how to compute the determinant of a matrix, and the incredible fact that we can perform expansion about *any* row *or* column to make this computation. In this largely theoretical section, we will state and prove several more intriguing properties about determinants. Our main goal will be the two results in Theorem SMZD [935] and Theorem DRMM [938], but more specifically, we will see how the value of a determinant will allow us to gain insight into the various properties of a square matrix.

Subsection DRO Determinants and Row Operations

We start easy with a straightforward theorem whose proof presages the style of subsequent proofs in this subsection.

Theorem DZRC Determinant with Zero Row or Column

Suppose that A is a square matrix with a row where every entry is zero, or a column where

every entry is zero. Then $\det(A) = 0$.

Proof Suppose that A is a square matrix of size n and row i has every entry equal to zero. We compute det (A) via expansion about row i.

$$\det (A) = \sum_{j=1}^{n} (-1)^{i+j} [A]_{ij} \det (A(i|j))$$
Theorem DER [897
$$= \sum_{j=1}^{n} (-1)^{i+j} 0 \det (A(i|j))$$
Row *i* is zeros
$$= \sum_{j=1}^{n} 0 = 0$$

The proof for the case of a zero column is entirely similar, or could be derived from an application of Theorem DT [899] employing the transpose of the matrix.

Theorem DRCS

Determinant for Row or Column Swap

Suppose that A is a square matrix. Let B be the square matrix obtained from A by interchanging the location of two rows, or interchanging the location of two columns. Then $\det(B) = -\det(A)$.

Proof Begin with the special case where A is a square matrix of size n and we form B by swapping *adjacent* rows i and i + 1 for some $1 \le i \le n - 1$. Notice that the assumption

about swapping adjacent rows means that B(i+1|j) = A(i|j) for all $1 \leq j \leq n$, and $[B]_{i+1,j} = [A]_{ij}$ for all $1 \leq j \leq n$. We compute det (B) via expansion about row i+1.

$$\det (B) = \sum_{j=1}^{n} (-1)^{(i+1)+j} [B]_{i+1,j} \det (B(i+1|j))$$
 Theorem DER [897]
$$= \sum_{j=1}^{n} (-1)^{(i+1)+j} [A]_{ij} \det (A(i|j))$$
 Hypothesis
$$= \sum_{j=1}^{n} (-1)^{1} (-1)^{i+j} [A]_{ij} \det (A(i|j))$$

$$= (-1) \sum_{j=1}^{n} (-1)^{i+j} [A]_{ij} \det (A(i|j))$$

$$= -\det (A)$$
 Theorem DER [897]

So the result holds for the special case where we swap adjacent rows of the matrix. As any computer scientist knows, we can accomplish *any* rearrangement of an ordered list by swapping adjacent elements. This principle can be demonstrated by naïve sorting algorithms such as "bubble sort." In any event, we don't need to discuss every possible reordering, we just need to consider a swap of two rows, say rows s and t with $1 \le s < t \le n$.

Begin with row s, and repeatedly swap it with each row just below it, including row t and stopping there. This will total t-s swaps. Now swap the former row t, which currently lives in row t-1, with each row above it, stopping when it becomes row s. This will total

another t - s - 1 swaps. In this way, we create B through a sequence of 2(t - s) - 1 swaps of adjacent rows, each of which adjusts det (A) by a multiplicative factor of -1. So

$$\det(B) = (-1)^{2(t-s)-1} \det(A) = ((-1)^2)^{t-s} (-1)^{-1} \det(A) = -\det(A)$$

as desired.

The proof for the case of swapping two columns is entirely similar, or could be derived from an application of Theorem DT [899] employing the transpose of the matrix.

So Theorem DRCS [920] tells us the effect of the first row operation (Definition RO [59]) on the determinant of a matrix. Here's the effect of the second row operation.

Theorem DRCM Determinant for Row or Column Multiples

Suppose that A is a square matrix. Let B be the square matrix obtained from A by multiplying a single row by the scalar α , or by multiplying a single column by the scalar α . Then det $(B) = \alpha \det(A)$.

Proof Suppose that A is a square matrix of size n and we form the square matrix B by multiplying each entry of row i of A by α . Notice that the other rows of A and B are equal, so A(i|j) = B(i|j), for all $1 \le j \le n$. We compute det (B) via expansion about row i.

$$\det(B) = \sum_{j=1}^{n} (-1)^{i+j} [B]_{ij} \det(B(i|j))$$
 Theorem DER [897]

$$= \sum_{j=1}^{n} (-1)^{i+j} [B]_{ij} \det (A(i|j))$$
Hypothesis
$$= \sum_{j=1}^{n} (-1)^{i+j} \alpha [A]_{ij} \det (A(i|j))$$
Hypothesis
$$= \alpha \sum_{j=1}^{n} (-1)^{i+j} [A]_{ij} \det (A(i|j))$$
$$= \alpha \det (A)$$
Theorem DER [897]

The proof for the case of a multiple of a column is entirely similar, or could be derived from an application of Theorem DT [899] employing the transpose of the matrix.

Let's go for understanding the effect of all three row operations. But first we need an intermediate result, but it is an easy one.

Theorem DERC

Determinant with Equal Rows or Columns

Suppose that A is a square matrix with two equal rows, or two equal columns. Then det(A) = 0.

Proof Suppose that A is a square matrix of size n where the two rows s and t are equal. Form the matrix B by swapping rows s and t. Notice that as a consequence of our hypothesis, A = B. Then

$$det (A) = \frac{1}{2} (det (A) + det (A))$$

= $\frac{1}{2} (det (A) - det (B))$
Theorem DRCS [920]
= $\frac{1}{2} (det (A) - det (A))$
Hypothesis, $A = B$
= $\frac{1}{2} (0) = 0$

The proof for the case of two equal columns is entirely similar, or could be derived from an application of Theorem DT [899] employing the transpose of the matrix.

Now explain the third row operation. Here we go.

Theorem DRCMA

Determinant for Row or Column Multiples and Addition

Suppose that A is a square matrix. Let B be the square matrix obtained from A by multiplying a row by the scalar α and then adding it to another row, or by multiplying a column by the scalar α and then adding it to another column. Then det $(B) = \det(A)$.

Proof Suppose that A is a square matrix of size n. Form the matrix B by multiplying row s by α and adding it to row t. Let C be the auxiliary matrix where we replace row t of A by row s of A. Notice that A(t|j) = B(t|j) = C(t|j) for all $1 \le j \le n$. We compute the

determinant of B by expansion about row t.

$$det (B) = \sum_{j=1}^{n} (-1)^{t+j} [B]_{tj} det (B(t|j))$$
Theorem DER [897]

$$= \sum_{j=1}^{n} (-1)^{t+j} \left(\alpha [A]_{sj} + [A]_{tj} \right) det (B(t|j))$$
Hypothesis

$$= \sum_{j=1}^{n} (-1)^{t+j} \alpha [A]_{sj} det (B(t|j))$$

$$+ \sum_{j=1}^{n} (-1)^{t+j} [A]_{tj} det (B(t|j))$$

$$= \alpha \sum_{j=1}^{n} (-1)^{t+j} [A]_{sj} det (B(t|j))$$

$$+ \sum_{j=1}^{n} (-1)^{t+j} [A]_{tj} det (B(t|j))$$

$$= \alpha \sum_{j=1}^{n} (-1)^{t+j} [A]_{tj} det (C(t|j))$$

$$+ \sum_{j=1}^{n} (-1)^{t+j} [A]_{tj} det (A(t|j))$$

 $= \alpha \det (C) + \det (A)$ Theorem DER [897] $= \alpha 0 + \det (A) = \det (A)$ Theorem DERC [923]

The proof for the case of adding a multiple of a column is entirely similar, or could be derived from an application of Theorem DT [899] employing the transpose of the matrix.

Is this what you expected? We could argue that the third row operation is the most popular, and yet it has no effect whatsoever on the determinant of a matrix! We can exploit this, along with our understanding of the other two row operations, to provide another approach to computing a determinant. We'll explain this in the context of an example.

Example DRO Determinant by row operations

Suppose we desire the determinant of the 4×4 matrix

$$A = \begin{bmatrix} 2 & 0 & 2 & 3 \\ 1 & 3 & -1 & 1 \\ -1 & 1 & -1 & 2 \\ 3 & 5 & 4 & 0 \end{bmatrix}$$

We will perform a sequence of row operations on this matrix, shooting for an upper triangular matrix, whose determinant will be simply the product of its diagonal entries. For each row operation, we will track the effect on the determinant via Theorem DRCS [920], Theorem

DRCM [922], Theorem DRCMA [924].

$$\begin{array}{c} \xrightarrow{R_1 \leftrightarrow R_2} A_1 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 2 & 0 & 2 & 3 \\ -1 & 1 & -1 & 2 \\ 3 & 5 & 4 & 0 \end{bmatrix} & \det(A) = -\det(A_1) & \text{Theorem DRCS [920]} \\ \end{array} \\ \begin{array}{c} \xrightarrow{-2R_1 + R_2} A_2 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & -6 & 4 & 1 \\ -1 & 1 & -1 & 2 \\ 3 & 5 & 4 & 0 \end{bmatrix} & = -\det(A_2) & \text{Theorem DRCMA [924]} \\ \end{array} \\ \begin{array}{c} \xrightarrow{1R_1 + R_3} A_3 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & -6 & 4 & 1 \\ 0 & 4 & -2 & 3 \\ 3 & 5 & 4 & 0 \end{bmatrix} & = -\det(A_3) & \text{Theorem DRCMA [924]} \\ \end{array} \\ \begin{array}{c} \xrightarrow{-3R_1 + R_4} A_4 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & -6 & 4 & 1 \\ 0 & 4 & -2 & 3 \\ 0 & -4 & 7 & -3 \end{bmatrix} & = -\det(A_4) & \text{Theorem DRCMA [924]} \\ \end{array} \\ \begin{array}{c} \xrightarrow{1R_3 + R_2} A_5 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & -2 & 2 & 4 \\ 0 & 4 & -2 & 3 \\ 0 & -4 & 7 & -3 \end{bmatrix} & = -\det(A_5) & \text{Theorem DRCMA [924]} \end{array}$$

$$\begin{array}{c} \stackrel{-\frac{1}{2}R_2}{\longrightarrow} A_6 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 4 & -2 & 3 \\ 0 & -4 & 7 & -3 \end{bmatrix} \\ \stackrel{-4R_2+R_3}{\longrightarrow} A_7 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 2 & 11 \\ 0 & -4 & 7 & -3 \end{bmatrix} \\ \stackrel{4R_2+R_4}{\longrightarrow} A_8 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 2 & 11 \\ 0 & 0 & 3 & -11 \end{bmatrix} \\ \stackrel{-1R_3+R_4}{\longrightarrow} A_9 = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 2 & 11 \\ 0 & 0 & 1 & -22 \end{bmatrix} \\ \stackrel{-2R_4+R_3}{\longrightarrow} A_{10} = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 55 \\ 0 & 0 & 1 & -22 \end{bmatrix}$$

$$= 2 \det (A_6)$$
 Theorem DRCM [922]
$$= 2 \det (A_7)$$
 Theorem DRCMA [924]
$$= 2 \det (A_8)$$
 Theorem DRCMA [924]
$$= 2 \det (A_9)$$
 Theorem DRCMA [924]
$$= 2 \det (A_{10})$$
 Theorem DRCMA [924]

 $\xrightarrow{R_{3} \leftrightarrow R_{4}} A_{11} = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & -22 \\ 0 & 0 & 0 & 55 \end{bmatrix} = -2 \det(A_{11}) \quad \text{Theorem DRCS [920]}$ $\xrightarrow{\frac{1}{55}R_{4}} A_{12} = \begin{bmatrix} 1 & 3 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 1 & -22 \\ 0 & 0 & 0 & 1 \end{bmatrix} = -110 \det(A_{12}) \quad \text{Theorem DRCM [922]}$

The matrix A_{12} is upper triangular, so expansion about the first column (repeatedly) will result in det $(A_{12}) = (1)(1)(1)(1) = 1$ (see Example DUTM [903]) and thus, det (A) = -110(1) = -110.

Notice that our sequence of row operations was somewhat *ad hoc*, such as the transformation to A_5 . We could have been even more methodical, and strictly followed the process that converts a matrix to reduced row-echelon form (Theorem REMEF [66]), eventually achieving the same numerical result with a final matrix that equaled the 4×4 identity matrix. Notice too that we could have stopped with A_8 , since at this point we could compute det (A_8) by two expansions about first columns, followed by a simple determinant of a 2×2 matrix (Theorem DMST [896]).

The beauty of this approach is that computationally we should already have written a procedure to convert matrices to reduced-row echelon form, so all we need to do is track the multiplicative changes to the determinant as the algorithm proceeds. Further, for a square

matrix of size n this approach requires on the order of n^3 multiplications, while a recursive application of expansion about a row or column (Theorem DER [897], Theorem DEC [901]) will require in the vicinity of (n - 1)(n!) multiplications. So even for very small matrices, a computational approach utilizing row operations will have superior run-time. Tracking, and controlling, the effects of round-off errors is another story, best saved for a numerical linear algebra course.

Subsection DROEM Determinants, Row Operations, Elementary Matrices

As a final preparation for our two most important theorems about determinants, we prove a handful of facts about the interplay of row operations and matrix multiplication with elementary matrices with regard to the determinant. But first, a simple, but crucial, fact about the identity matrix.

Theorem DIM Determinant of the Identity Matrix For every $n \ge 1$, det $(I_n) = 1$.

Proof It may be overkill, but this is a good situation to run through a proof by induction on n (Technique I [1630]). Is the result true when n = 1? Yes,

 $\det (I_1) = [I_1]_{11} \qquad \qquad \text{Definition DM [894]}$

= 1 Definition IM [175]

Now assume the theorem is true for the identity matrix of size n - 1 and investigate the determinant of the identity matrix of size n with expansion about row 1,

$$\det (I_n) = \sum_{j=1}^n (-1)^{1+j} [I_n]_{1j} \det (I_n (1|j))$$
 Definition DM [894]
$$= (-1)^{1+1} [I_n]_{11} \det (I_n (1|1)) + \sum_{j=2}^n (-1)^{1+j} [I_n]_{1j} \det (I_n (1|j))$$

$$= 1 \det (I_{n-1}) + \sum_{j=2}^n (-1)^{1+j} 0 \det (I_n (1|j))$$
 Definition IM [175]
$$= 1(1) + \sum_{j=2}^n 0 = 1$$
 Induction Hypothesis

Theorem DEM

Determinants of Elementary Matrices

For the three possible versions of an elementary matrix (Definition ELEM [883]) we have the determinants,

- 1. det $(E_{i,j}) = -1$
- 2. det $(E_i(\alpha)) = \alpha$
- 3. det $(E_{i,j}(\alpha)) = 1$

Proof Swapping rows *i* and *j* of the identity matrix will create $E_{i,j}$ (Definition ELEM [883]), so

$$det (E_{i,j}) = -det (I_n)$$

$$= -1$$
Theorem DRCS [920]
Theorem DIM [930]

Multiplying row i of the identity matrix by α will create $E_i(\alpha)$ (Definition ELEM [883]), so

$\det\left(E_{i}\left(\alpha\right)\right) = \alpha \det\left(I_{n}\right)$	Theorem DRCM [922]
$= \alpha(1) = \alpha$	Theorem DIM [930]

Multiplying row *i* of the identity matrix by α and adding to row *j* will create $E_{i,j}(\alpha)$ (Definition ELEM [883]), so

 $det (E_{i,j} (\alpha)) = det (I_n)$ = 1Theorem DRCMA [924]
Theorem DIM [930]

Theorem DEMMM Determinants, Elementary Matrices, Matrix Multiplication

Suppose that A is a square matrix of size n and E is any elementary matrix of size n. Then

 $\det (EA) = \det (E) \det (A)$

Proof The proof procedes in three parts, one for each type of elementary matrix, with each part very similar to the other two. First, let B be the matrix obtained from A by swapping rows i and j,

$$det (E_{i,j}A) = det (B)$$

$$= - det (A)$$

$$= det (E_{i,j}) det (A)$$
Theorem DRCS [920]
Theorem DEM [932]

 \square

Second, let B be the matrix obtained from A by multiplying row i by α ,

$$det (E_i (\alpha) A) = det (B)$$

$$= \alpha det (A)$$

$$= det (E_i (\alpha)) det (A)$$
Theorem DEM [932]
Theorem DEM [932]

Third, let B be the matrix obtained from A by multiplying row i by α and adding to row j,

 $det (E_{i,j} (\alpha) A) = det (B)$ = det (A) $= det (E_{i,j} (\alpha)) det (A)$ Theorem DRCMA [924]
Theorem DEM [932]

Since the desired result holds for each variety of elementary matrix individually, we are done.

Subsection DNMMM Determinants, Nonsingular Matrices, Matrix Multiplication

If you asked someone with substantial experience working with matrices about the value of the determinant, they'd be likely to quote the following theorem as the first thing to come to mind.

Theorem SMZD

Singular Matrices have Zero Determinants

Let A be a square matrix. Then A is singular if and only if det (A) = 0.

Proof Rather than jumping into the two halves of the equivalence, we first establish a few items. Let B be the unique square matrix that is row-equivalent to A and in reduced row-echelon form (Theorem REMEF [66], Theorem RREFU [70]). For each of the row operations that converts B into A, there is an elementary matrix E_i which effects the row operation by matrix multiplication (Theorem EMDRO [886]). Repeated applications of Theorem EMDRO [886] allow us to write

$$A = E_s E_{s-1} \dots E_2 E_1 B$$

Then

$$\det (A) = \det (E_s E_{s-1} \dots E_2 E_1 B)$$

= det $(E_s) \det (E_{s-1}) \dots \det (E_2) \det (E_1) \det (B)$ Theorem DEMMM [933]

From Theorem DEM [932] we can infer that the determinant of an elementary matrix is never zero (note the ban on $\alpha = 0$ for $E_i(\alpha)$ in Definition ELEM [883]). So the product on the right is composed of nonzero scalars, with the possible exception of det (B). More precisely, we can argue that det (A) = 0 if and only if det (B) = 0. With this established, we can take up the two halves of the equivalence.

 (\Rightarrow) If A is singular, then by Theorem NMRRI [176], B cannot be the identity matrix. Because (1) the number of pivot columns is equal to the number of nonzero rows, (2) not every column is a pivot column, and (3) B is square, we see that B must have a zero row.

 \square

By Theorem DZRC [919] the determinant of B is zero, and by the above, we conclude that the determinant of A is zero.

(\Leftarrow) We will prove the contrapositive (Technique CP [1621]). So assume A is nonsingular, then by Theorem NMRRI [176], B is the identity matrix and Theorem DIM [930] tells us that det (B) = 1 \neq 0. With the argument above, we conclude that the determinant of A is nonzero as well.

For the case of 2×2 matrices you might compare the application of Theorem SMZD [935] with the combination of the results stated in Theorem DMST [896] and Theorem TTMI [509].

Example ZNDAB

Zero and nonzero determinant, Archetypes A and B

The coefficient matrix in Archetype A [1642] has a zero determinant (check this!) while the coefficient matrix Archetype B [1652] has a nonzero determinant (check this, too). These matrices are singular and nonsingular, respectively. This is exactly what Theorem SMZD [935] says, and continues our list of contrasts between these two archetypes. \boxtimes

Since Theorem SMZD [935] is an equivalence (Technique E [1619]) we can expand on our growing list of equivalences about nonsingular matrices. The addition of the condition det $(A) \neq 0$ is one of the best motivations for learning about determinants.

Theorem NME7

Nonsingular Matrix Equivalences, Round 7

Suppose that A is a square matrix of size n. The following are equivalent.

1. A is nonsingular.

- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.
- 7. The column space of A is \mathbb{C}^n , $\mathcal{C}(A) = \mathbb{C}^n$.
- 8. The columns of A are a basis for \mathbb{C}^n .
- 9. The rank of A is n, r(A) = n.
- 10. The nullity of A is zero, n(A) = 0.
- 11. The determinant of A is nonzero, det $(A) \neq 0$.

Proof Theorem SMZD [935] says A is singular if and only if det (A) = 0. If we negate each of these statements, we arrive at two contrapositives that we can combine as the equivalence, A is nonsingular if and only if det $(A) \neq 0$. This allows us to add a new statement to the list found in Theorem NME6 [834].

Computationally, row-reducing a matrix is the most efficient way to determine if a matrix is nonsingular, though the effect of using division in a computer can lead to round-off errors

 \square

that confuse small quantities with critical zero quantities. Conceptually, the determinant may seem the most efficient way to determine if a matrix is nonsingular. The definition of a determinant uses just addition, subtraction and multiplication, so division is never a problem. And the final test is easy: is the determinant zero or not? However, the number of operations involved in computing a determinant by the definition very quickly becomes so excessive as to be impractical.

Now for the *coup de grâce*. We will generalize Theorem DEMMM [933] to the case of *any* two square matrices. You may recall thinking that matrix multiplication was defined in a needlessly complicated manner. For sure, the definition of a determinant seems even stranger. (Though Theorem SMZD [935] might be forcing you to reconsider.) Read the statement of the next theorem and contemplate how nicely matrix multiplication and determinants play with each other.

Theorem DRMM

Determinant Respects Matrix Multiplication

Suppose that A and B are square matrices of the same size. Then det (AB) = det (A) det (B).

Proof This proof is constructed in two cases. First, suppose that A is singular. Then det (A) = 0 by Theorem SMZD [935]. By the contrapositive of Theorem NPNT [538], AB is singular as well. So by a second application of Theorem SMZD [935], det (AB) = 0. Putting it all together

$$\det (AB) = 0 = 0 \det (B) = \det (A) \det (B)$$

as desired.

For the second case, suppose that A is nonsingular. By Theorem NMPEM [892] there are elementary matrices $E_1, E_2, E_3, \ldots, E_s$ such that $A = E_1 E_2 E_3 \ldots E_s$. Then

$$det (AB) = det (E_1 E_2 E_3 \dots E_s B)$$

= det (E₁) det (E₂) det (E₃) ... det (E_s) det (B) Theorem DEMMM [933]
= det (E_1 E_2 E_3 \dots E_s) det (B) Theorem DEMMM [933]
= det (A) det (B)

It is amazing that matrix multiplication and the determinant interact this way. Might it also be true that $\det(A + B) = \det(A) + \det(B)$? (See Exercise PDM.M30 [941].)

Subsection READ Reading Questions

1. Consider the two matrices below, and suppose you already have computed det (A) = -120. What is det (B)? Why?

$$A = \begin{bmatrix} 0 & 8 & 3 & -4 \\ -1 & 2 & -2 & 5 \\ -2 & 8 & 4 & 3 \\ 0 & -4 & 2 & -3 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 0 & 8 & 3 & -4 \\ 0 & -4 & 2 & -3 \\ -2 & 8 & 4 & 3 \\ -1 & 2 & -2 & 5 \end{bmatrix}$$

- 2. State the theorem that allows us to make yet another extension to our NMEx series of theorems.
- 3. What is amazing about the interaction between matrix multiplication and the determinant?

Subsection EXC Exercises

C30 Each of the archetypes below is a system of equations with a square coefficient matrix, or is a square matrix itself. Compute the determinant of each matrix, noting how Theorem SMZD [935] indicates when the matrix is singular or nonsingular.

Archetype A [1642] Archetype B [1652] Archetype F [1686] Archetype K [1734] Archetype L [1742]

Contributed by Robert Beezer

M20 Construct a 3×3 nonsingular matrix and call it A. Then, for each entry of the matrix, compute the corresponding cofactor, and create a new 3×3 matrix full of these cofactors by placing the cofactor of an entry in the same location as the entry it was based on. Once complete, call this matrix C. Compute AC^t . Any observations? Repeat with a new matrix, or perhaps with a 4×4 matrix. Contributed by Robert Beezer Solution [943]

M30 Construct an example to show that the following statement is not true for all square

matrices A and B of the same size: $\det(A + B) = \det(A) + \det(B)$. Contributed by Robert Beezer

T10 Theorem NPNT [538] says that if the product of square matrices AB is nonsingular, then the individual matrices A and B are nonsingular also. Construct a new proof of this result making use of theorems about determinants of matrices.

Contributed by Robert Beezer

T15 Use Theorem DRCM [922] to prove Theorem DZRC [919] as a corollary. (See Technique LC [1634].) Contributed by Robert Beezer

T20 Suppose that A is a square matrix of size n and $\alpha \in \mathbb{C}$ is a scalar. Prove that $\det(\alpha A) = \alpha^n \det(A)$. Contributed by Robert Beezer

T25 Employ Theorem DT [899] to construct the second half of the proof of Theorem DRCM [922] (the portion about a multiple of a column). Contributed by Robert Beezer

Subsection SOL Solutions

M20 Contributed by Robert Beezer Statement [941]

The result of these computations should be a matrix with the value of det (A) in the diagonal entries and zeros elsewhere. The suggestion of using a nonsingular matrix was partially so that it was obvious that the value of the determinant appears on the diagonal.

This result (which is true in general) provides a method for computing the inverse of a nonsingular matrix. Since $AC^t = \det(A) I_n$, we can multiply by the reciprocal of the determinant (which is nonzero!) and the inverse of A (it exists!) to arrive at an expression for the matrix inverse:

$$A^{-1} = \frac{1}{\det\left(A\right)}C^{t}$$

Chapter E Eigenvalues

When we have a square matrix of size n, A, and we multiply it by a vector \mathbf{x} from \mathbb{C}^n to form the matrix-vector product (Definition MVP [461]), the result is another vector in \mathbb{C}^n . So we can adopt a functional view of this computation — the act of multiplying by a square matrix is a function that converts one vector (\mathbf{x}) into another one ($A\mathbf{x}$) of the same size. For some vectors, this seemingly complicated computation is really no more complicated than scalar multiplication. The vectors vary according to the choice of A, so the question is to determine, for an individual choice of A, if there are any such vectors, and if so, which ones. It happens in a variety of situations that these vectors (and the scalars that go along with them) are of special interest.

We will be solving polynomial equations in this chapter, which raises the specter of roots

that are complex numbers. This distinct possibility is our main reason for entertaining the complex numbers throughout the course. You might be moved to revisit Section CNO [1594] and Section O [395].

Section EE Eigenvalues and Eigenvectors

We start with the principal definition for this chapter.

Subsection EEM Eigenvalues and Eigenvectors of a Matrix

Definition EEM Eigenvalues and Eigenvectors of a Matrix

Suppose that A is a square matrix of size $n, \mathbf{x} \neq \mathbf{0}$ is a vector in \mathbb{C}^n , and λ is a scalar in \mathbb{C} . Then we say \mathbf{x} is an **eigenvector** of A with **eigenvalue** λ if

 $A\mathbf{x} = \lambda \mathbf{x}$

 \triangle

Before going any further, perhaps we should convince you that such things ever happen at all. Understand the next example, but do not concern yourself with where the pieces come from. We will have methods soon enough to be able to discover these eigenvectors ourselves.

Example SEE

Some eigenvalues and eigenvectors

Consider the matrix

$$A = \begin{bmatrix} 204 & 98 & -26 & -10 \\ -280 & -134 & 36 & 14 \\ 716 & 348 & -90 & -36 \\ -472 & -232 & 60 & 28 \end{bmatrix}$$

and the vectors

$$\mathbf{x} = \begin{bmatrix} 1\\ -1\\ 2\\ 5 \end{bmatrix} \qquad \qquad \mathbf{y} = \begin{bmatrix} -3\\ 4\\ -10\\ 4 \end{bmatrix} \qquad \qquad \mathbf{z} = \begin{bmatrix} -3\\ 7\\ 0\\ 8 \end{bmatrix} \qquad \qquad \mathbf{w} = \begin{bmatrix} 1\\ -1\\ 4\\ 0 \end{bmatrix}$$

Then

$$A\mathbf{x} = \begin{bmatrix} 204 & 98 & -26 & -10\\ -280 & -134 & 36 & 14\\ 716 & 348 & -90 & -36\\ -472 & -232 & 60 & 28 \end{bmatrix} \begin{bmatrix} 1\\ -1\\ 2\\ 5 \end{bmatrix} = \begin{bmatrix} 4\\ -4\\ 8\\ 20 \end{bmatrix} = 4 \begin{bmatrix} 1\\ -1\\ 2\\ 5 \end{bmatrix} = 4\mathbf{x}$$

so **x** is an eigenvector of A with eigenvalue $\lambda = 4$. Also,

$$A\mathbf{y} = \begin{bmatrix} 204 & 98 & -26 & -10\\ -280 & -134 & 36 & 14\\ 716 & 348 & -90 & -36\\ -472 & -232 & 60 & 28 \end{bmatrix} \begin{bmatrix} -3\\4\\-10\\4 \end{bmatrix} = \begin{bmatrix} 0\\0\\0\\0 \end{bmatrix} = 0 \begin{bmatrix} -3\\4\\-10\\4 \end{bmatrix} = 0\mathbf{y}$$

so **y** is an eigenvector of A with eigenvalue $\lambda = 0$. Also,

$$A\mathbf{z} = \begin{bmatrix} 204 & 98 & -26 & -10\\ -280 & -134 & 36 & 14\\ 716 & 348 & -90 & -36\\ -472 & -232 & 60 & 28 \end{bmatrix} \begin{bmatrix} -3\\7\\0\\8 \end{bmatrix} = \begin{bmatrix} -6\\14\\0\\16 \end{bmatrix} = 2\begin{bmatrix} -3\\7\\0\\8 \end{bmatrix} = 2\mathbf{z}$$

so **z** is an eigenvector of A with eigenvalue $\lambda = 2$. Also,

$$A\mathbf{w} = \begin{bmatrix} 204 & 98 & -26 & -10\\ -280 & -134 & 36 & 14\\ 716 & 348 & -90 & -36\\ -472 & -232 & 60 & 28 \end{bmatrix} \begin{bmatrix} 1\\ -1\\ 4\\ 0 \end{bmatrix} = \begin{bmatrix} 2\\ -2\\ 8\\ 0 \end{bmatrix} = 2\begin{bmatrix} 1\\ -1\\ 4\\ 0 \end{bmatrix} = 2\mathbf{w}$$

so **w** is an eigenvector of A with eigenvalue $\lambda = 2$.

So we have demonstrated four eigenvectors of A. Are there more? Yes, any nonzero scalar multiple of an eigenvector is again an eigenvector. In this example, set $\mathbf{u} = 30\mathbf{x}$. Then

$$A\mathbf{u} = A(30\mathbf{x})$$

$= 30A\mathbf{x}$	Theorem MMSMM [477]
$= 30(4\mathbf{x})$	${\bf x}$ an eigenvector of A
$=4(30\mathbf{x})$	Property SMAM [431]
$=4\mathbf{u}$	

so that **u** is also an eigenvector of A for the same eigenvalue, $\lambda = 4$.

The vectors \mathbf{z} and \mathbf{w} are both eigenvectors of A for the same eigenvalue $\lambda = 2$, yet this is not as simple as the two vectors just being scalar multiples of each other (they aren't). Look what happens when we add them together, to form $\mathbf{v} = \mathbf{z} + \mathbf{w}$, and multiply by A,

$A\mathbf{v} = A(\mathbf{z} + \mathbf{w})$	
$= A\mathbf{z} + A\mathbf{w}$	Theorem MMDAA [476]
$=2\mathbf{z}+2\mathbf{w}$	\mathbf{z}, \mathbf{w} eigenvectors of A
$=2(\mathbf{z}+\mathbf{w})$	Property DVAC [206]
$= 2\mathbf{v}$	

so that **v** is also an eigenvector of A for the eigenvalue $\lambda = 2$. So it would appear that the set of eigenvectors that are associated with a fixed eigenvalue is closed under the vector space operations of \mathbb{C}^n . Hmmm.

The vector \mathbf{y} is an eigenvector of A for the eigenvalue $\lambda = 0$, so we can use Theorem ZSSM [676] to write $A\mathbf{y} = 0\mathbf{y} = \mathbf{0}$. But this also means that $\mathbf{y} \in \mathcal{N}(A)$. There would appear to be a connection here also.

Example SEE [946] hints at a number of intriguing properties, and there are many more.

We will explore the general properties of eigenvalues and eigenvectors in Section PEE [1001], but in this section we will concern ourselves with the question of actually computing eigenvalues and eigenvectors. First we need a bit of background material on polynomials and matrices.

Subsection PM Polynomials and Matrices

A polynomial is a combination of powers, multiplication by scalar coefficients, and addition (with subtraction just being the inverse of addition). We never have occasion to divide when computing the value of a polynomial. So it is with matrices. We can add and subtract matrices, we can multiply matrices by scalars, and we can form powers of square matrices by repeated applications of matrix multiplication. We do not normally divide matrices (though sometimes we can multiply by an inverse). If a matrix is square, all the operations constituting a polynomial will preserve the size of the matrix. So it is natural to consider evaluating a polynomial with a matrix, effectively replacing the variable of the polynomial by a matrix. We'll demonstrate with an example,

Example PM Polynomial of a matrix Let

$$p(x) = 14 + 19x - 3x^2 - 7x^3 + x^4 \qquad D = \begin{bmatrix} -1 & 3 & 2\\ 1 & 0 & -2\\ -3 & 1 & 1 \end{bmatrix}$$

and we will compute p(D). First, the necessary powers of D. Notice that D^0 is defined to be the multiplicative identity, I_3 , as will be the case in general.

$$D^{0} = I_{3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$D^{1} = D = \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix}$$

$$D^{2} = DD^{1} = \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -1 & -6 \\ 5 & 1 & 0 \\ 1 & -8 & -7 \end{bmatrix}$$

$$D^{3} = DD^{2} = \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & -1 & -6 \\ 5 & 1 & 0 \\ 1 & -8 & -7 \end{bmatrix} = \begin{bmatrix} 19 & -12 & -8 \\ -4 & 15 & 8 \\ 12 & -4 & 11 \end{bmatrix}$$

$$D^{4} = DD^{3} = \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 19 & -12 & -8 \\ -4 & 15 & 8 \\ 12 & -4 & 11 \end{bmatrix} = \begin{bmatrix} -7 & 49 & 54 \\ -5 & -4 & -30 \\ -49 & 47 & 43 \end{bmatrix}$$

Then

$$\begin{split} p(D) &= 14 + 19D - 3D^2 - 7D^3 + D^4 \\ &= 14 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + 19 \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} - 3 \begin{bmatrix} -2 & -1 & -6 \\ 5 & 1 & 0 \\ 1 & -8 & -7 \end{bmatrix} \\ &- 7 \begin{bmatrix} 19 & -12 & -8 \\ -4 & 15 & 8 \\ 12 & -4 & 11 \end{bmatrix} + \begin{bmatrix} -7 & 49 & 54 \\ -5 & -4 & -30 \\ -49 & 47 & 43 \end{bmatrix} \\ &= \begin{bmatrix} -139 & 193 & 166 \\ 27 & -98 & -124 \\ -193 & 118 & 20 \end{bmatrix} \end{split}$$

Notice that p(x) factors as

$$p(x) = 14 + 19x - 3x^{2} - 7x^{3} + x^{4} = (x - 2)(x - 7)(x + 1)^{2}$$

Because D commutes with itself (DD = DD), we can use distributivity of matrix multiplication across matrix addition (Theorem MMDAA [476]) without being careful with any of the matrix products, and just as easily evaluate p(D) using the factored form of p(x),

$$p(D) = 14 + 19D - 3D^2 - 7D^3 + D^4 = (D - 2I_3)(D - 7I_3)(D + I_3)^2$$

Version 2.99

$$= \begin{bmatrix} -3 & 3 & 2\\ 1 & -2 & -2\\ -3 & 1 & -1 \end{bmatrix} \begin{bmatrix} -8 & 3 & 2\\ 1 & -7 & -2\\ -3 & 1 & -6 \end{bmatrix} \begin{bmatrix} 0 & 3 & 2\\ 1 & 1 & -2\\ -3 & 1 & 2 \end{bmatrix}^{2}$$
$$= \begin{bmatrix} -139 & 193 & 166\\ 27 & -98 & -124\\ -193 & 118 & 20 \end{bmatrix}$$

This example is not meant to be too profound. It is meant to show you that it is natural to evaluate a polynomial with a matrix, and that the factored form of the polynomial is as good as (or maybe better than) the expanded form. And do not forget that constant terms in polynomials are really multiples of the identity matrix when we are evaluating the polynomial with a matrix. \boxtimes

Subsection EEE Existence of Eigenvalues and Eigenvectors

Before we embark on computing eigenvalues and eigenvectors, we will prove that every matrix has at least one eigenvalue (and an eigenvector to go with it). Later, in Theorem MNEM [1019], we will determine the maximum number of eigenvalues a matrix may have.

The determinant (Definition DM [894]) will be a powerful tool in Subsection EE.CEE [960] when it comes time to compute eigenvalues. However, it is possible, with some more

advanced machinery, to compute eigenvalues without ever making use of the determinant. Sheldon Axler does just that in his book, *Linear Algebra Done Right*. Here and now, we give Axler's "determinant-free" proof that every matrix has an eigenvalue. The result is not too startling, but the proof is most enjoyable.

Theorem EMHE

Every Matrix Has an Eigenvalue

Suppose A is a square matrix. Then A has at least one eigenvalue.

Proof Suppose that A has size n, and choose **x** as any nonzero vector from \mathbb{C}^n . (Notice how much latitude we have in our choice of **x**. Only the zero vector is off-limits.) Consider the set

$$S = \left\{ \mathbf{x}, \, A\mathbf{x}, \, A^2\mathbf{x}, \, A^3\mathbf{x}, \, \dots, \, A^n\mathbf{x} \right\}$$

This is a set of n+1 vectors from \mathbb{C}^n , so by Theorem MVSLD [328], S is linearly dependent. Let $a_0, a_1, a_2, \ldots, a_n$ be a collection of n+1 scalars from \mathbb{C} , not all zero, that provide a relation of linear dependence on S. In other words,

$$a_0\mathbf{x} + a_1A\mathbf{x} + a_2A^2\mathbf{x} + a_3A^3\mathbf{x} + \dots + a_nA^n\mathbf{x} = \mathbf{0}$$

Some of the a_i are nonzero. Suppose that just $a_0 \neq 0$, and $a_1 = a_2 = a_3 = \cdots = a_n = 0$. Then $a_0 \mathbf{x} = \mathbf{0}$ and by Theorem SMEZV [679], either $a_0 = 0$ or $\mathbf{x} = \mathbf{0}$, which are both contradictions. So $a_i \neq 0$ for some $i \geq 1$. Let m be the largest integer such that $a_m \neq 0$. From this discussion we know that $m \geq 1$. We can also assume that $a_m = 1$, for if not, replace each a_i by a_i/a_m to obtain scalars that serve equally well in providing a relation of linear dependence on S.

Define the polynomial

$$p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots + a_m x^m$$

Because we have consistently used \mathbb{C} as our set of scalars (rather than \mathbb{R}), we know that we can factor p(x) into linear factors of the form $(x - b_i)$, where $b_i \in \mathbb{C}$. So there are scalars, $b_1, b_2, b_3, \ldots, b_m$, from \mathbb{C} so that,

$$p(x) = (x - b_m)(x - b_{m-1}) \cdots (x - b_3)(x - b_2)(x - b_1)$$

Put it all together and

$$\mathbf{0} = a_0 \mathbf{x} + a_1 A \mathbf{x} + a_2 A^2 \mathbf{x} + a_3 A^3 \mathbf{x} + \dots + a_n A^n \mathbf{x}$$

$$= a_0 \mathbf{x} + a_1 A \mathbf{x} + a_2 A^2 \mathbf{x} + a_3 A^3 \mathbf{x} + \dots + a_m A^m \mathbf{x}$$

$$= (a_0 I_n + a_1 A + a_2 A^2 + a_3 A^3 + \dots + a_m A^m) \mathbf{x}$$

$$= p(A) \mathbf{x}$$

$$= (A - b_m I_n)(A - b_{m-1} I_n) \cdots (A - b_3 I_n)(A - b_2 I_n)(A - b_1 I_n) \mathbf{x}$$

Theorem MMDAA [476]
Definition of $p(x)$

Let k be the smallest integer such that

$$(A - b_k I_n)(A - b_{k-1} I_n) \cdots (A - b_3 I_n)(A - b_2 I_n)(A - b_1 I_n) \mathbf{x} = \mathbf{0}.$$

From the preceding equation, we know that $k \leq m$. Define the vector **z** by

$$\mathbf{z} = (A - b_{k-1}I_n) \cdots (A - b_3I_n)(A - b_2I_n)(A - b_1I_n)\mathbf{x}$$

Notice that by the definition of k, the vector \mathbf{z} must be nonzero. In the case where k = 1, we understand that \mathbf{z} is defined by $\mathbf{z} = \mathbf{x}$, and \mathbf{z} is still nonzero. Now

$$(A - b_k I_n)\mathbf{z} = (A - b_k I_n)(A - b_{k-1} I_n) \cdots (A - b_3 I_n)(A - b_2 I_n)(A - b_1 I_n)\mathbf{x} = \mathbf{0}$$

which allows us to write

$A\mathbf{z} = (A + \mathcal{O})\mathbf{z}$	Property ZM $[431]$
$= (A - b_k I_n + b_k I_n) \mathbf{z}$	Property AIM [431]
$= (A - b_k I_n)\mathbf{z} + b_k I_n \mathbf{z}$	Theorem MMDAA [476]
$= 0 + b_k I_n \mathbf{z}$	Defining property of ${\bf z}$
$= b_k I_n \mathbf{z}$	Property ZM [431]
$= b_k \mathbf{z}$	Theorem MMIM [475]

Since $\mathbf{z} \neq \mathbf{0}$, this equation says that \mathbf{z} is an eigenvector of A for the eigenvalue $\lambda = b_k$ (Definition EEM [945]), so we have shown that any square matrix A does have at least one eigenvalue.

The proof of Theorem EMHE [953] is constructive (it contains an unambiguous procedure that leads to an eigenvalue), but it is not meant to be practical. We will illustrate the theorem with an example, the purpose being to provide a companion for studying the proof and not to suggest this is the best procedure for computing an eigenvalue.

Example CAEHW

Computing an eigenvalue the hard way

This example illustrates the proof of Theorem EMHE [953], so will employ the same notation

as the proof — look there for full explanations. It is *not* meant to be an example of a reasonable computational approach to finding eigenvalues and eigenvectors. OK, warnings in place, here we go.

Let

$$A = \begin{bmatrix} -7 & -1 & 11 & 0 & -4 \\ 4 & 1 & 0 & 2 & 0 \\ -10 & -1 & 14 & 0 & -4 \\ 8 & 2 & -15 & -1 & 5 \\ -10 & -1 & 16 & 0 & -6 \end{bmatrix}$$

and choose

 $\mathbf{x} = \begin{bmatrix} 3\\0\\3\\-5\\4 \end{bmatrix}$

It is important to notice that the choice of \mathbf{x} could be *anything*, so long as it is *not* the zero vector. We have not chosen \mathbf{x} totally at random, but so as to make our illustration of the theorem as general as possible. You could replicate this example with your own choice and the computations are guaranteed to be reasonable, provided you have a computational tool that will factor a fifth degree polynomial for you.

The set

$$S = \left\{ \mathbf{x}, \, A\mathbf{x}, \, A^2\mathbf{x}, \, A^3\mathbf{x}, \, A^4\mathbf{x}, \, A^5\mathbf{x} \right\}$$

$$= \left\{ \begin{bmatrix} 3\\0\\3\\-5\\4 \end{bmatrix}, \begin{bmatrix} -4\\2\\-4\\4\\-6 \end{bmatrix}, \begin{bmatrix} 6\\-6\\6\\-2\\10 \end{bmatrix}, \begin{bmatrix} -10\\14\\-10\\-2\\-18 \end{bmatrix}, \begin{bmatrix} 18\\-30\\18\\10\\34 \end{bmatrix}, \begin{bmatrix} -34\\62\\-34\\-26\\-66 \end{bmatrix} \right\}$$

is guaranteed to be linearly dependent, as it has six vectors from \mathbb{C}^5 (Theorem MVSLD [328]). We will search for a non-trivial relation of linear dependence by solving a homogeneous system of equations whose coefficient matrix has the vectors of S as columns through row operations,

There are four free variables for describing solutions to this homogeneous system, so we have our pick of solutions. The most expedient choice would be to set $x_3 = 1$ and $x_4 = x_5 = x_6 = 0$. However, we will again opt to maximize the generality of our illustration of Theorem EMHE [953] and choose $x_3 = -8$, $x_4 = -3$, $x_5 = 1$ and $x_6 = 0$. The leads to a solution with $x_1 = 16$ and $x_2 = 12$.

This relation of linear dependence then says that

$$0 = 16x + 12Ax - 8A^{2}x - 3A^{3}x + A^{4}x + 0A^{5}x$$

$$\mathbf{0} = (16 + 12A - 8A^2 - 3A^3 + A^4) \mathbf{x}$$

So we define $p(x) = 16 + 12x - 8x^2 - 3x^3 + x^4$, and as advertised in the proof of Theorem EMHE [953], we have a polynomial of degree m = 4 > 1 such that $p(A)\mathbf{x} = \mathbf{0}$. Now we need to factor p(x) over \mathbb{C} . If you made your own choice of \mathbf{x} at the start, this is where you might have a fifth degree polynomial, and where you might need to use a computational tool to find roots and factors. We have

$$p(x) = 16 + 12x - 8x^2 - 3x^3 + x^4 = (x - 4)(x + 2)(x - 2)(x + 1)$$

So we know that

$$\mathbf{0} = p(A)\mathbf{x} = (A - 4I_5)(A + 2I_5)(A - 2I_5)(A + 1I_5)\mathbf{x}$$

We apply one factor at a time, until we get the zero vector, so as to determine the value of k described in the proof of Theorem EMHE [953],

$$(A+1I_5)\mathbf{x} = \begin{bmatrix} -6 & -1 & 11 & 0 & -4 \\ 4 & 2 & 0 & 2 & 0 \\ -10 & -1 & 15 & 0 & -4 \\ 8 & 2 & -15 & 0 & 5 \\ -10 & -1 & 16 & 0 & -5 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 3 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \\ -1 \\ -2 \end{bmatrix}$$
$$(A-2I_5)(A+1I_5)\mathbf{x} = \begin{bmatrix} -9 & -1 & 11 & 0 & -4 \\ 4 & -1 & 0 & 2 & 0 \\ -10 & -1 & 12 & 0 & -4 \\ 8 & 2 & -15 & -3 & 5 \\ -10 & -1 & 16 & 0 & -8 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -1 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ -8 \\ 4 \\ 4 \\ 8 \end{bmatrix}$$

$$(A+2I_5)(A-2I_5)(A+1I_5)\mathbf{x} = \begin{bmatrix} -5 & -1 & 11 & 0 & -4\\ 4 & 3 & 0 & 2 & 0\\ -10 & -1 & 16 & 0 & -4\\ 8 & 2 & -15 & 1 & 5\\ -10 & -1 & 16 & 0 & -4 \end{bmatrix} \begin{bmatrix} 4\\ -8\\ 4\\ 4\\ 8 \end{bmatrix} = \begin{bmatrix} 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix}$$

So k = 3 and

$$\mathbf{z} = (A - 2I_5)(A + 1I_5)\mathbf{x} = \begin{bmatrix} 4\\-8\\4\\4\\8 \end{bmatrix}$$

is an eigenvector of A for the eigenvalue $\lambda = -2$, as you can check by doing the computation Az. If you work through this example with your own choice of the vector **x** (strongly recommended) then the eigenvalue you will find may be different, but will be in the set $\{3, 0, 1, -1, -2\}$. See Exercise EE.M60 [987] for a suggested starting vector.

Subsection CEE Computing Eigenvalues and Eigenvectors

Fortunately, we need not rely on the procedure of Theorem EMHE [953] each time we need an eigenvalue. It is the determinant, and specifically Theorem SMZD [935], that provides the main tool for computing eigenvalues. Here is an informal sequence of equivalences that is the key to determining the eigenvalues and eigenvectors of a matrix,

$$A\mathbf{x} = \lambda \mathbf{x} \iff A\mathbf{x} - \lambda I_n \mathbf{x} = \mathbf{0} \iff (A - \lambda I_n) \mathbf{x} = \mathbf{0}$$

So, for an eigenvalue λ and associated eigenvector $\mathbf{x} \neq \mathbf{0}$, the vector \mathbf{x} will be a nonzero element of the null space of $A - \lambda I_n$, while the matrix $A - \lambda I_n$ will be singular and therefore have zero determinant. These ideas are made precise in Theorem EMRCP [962] and Theorem EMNS [965], but for now this brief discussion should suffice as motivation for the following definition and example.

Definition CP Characteristic Polynomial

Suppose that A is a square matrix of size n. Then the **characteristic polynomial** of A is the polynomial $p_A(x)$ defined by

$$p_A(x) = \det\left(A - xI_n\right)$$

Example CPMS3 Characteristic polynomial of a matrix, size 3 Consider

$$F = \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix}$$

Then

$$p_{F}(x) = \det (F - xI_{3})$$

$$= \begin{vmatrix} -13 - x & -8 & -4 \\ 12 & 7 - x & 4 \\ 24 & 16 & 7 - x \end{vmatrix}$$
Definition CP [960]
$$= (-13 - x) \begin{vmatrix} 7 - x & 4 \\ 16 & 7 - x \end{vmatrix} + (-8)(-1) \begin{vmatrix} 12 & 4 \\ 24 & 7 - x \end{vmatrix}$$
Definition DM [894]
$$+ (-4) \begin{vmatrix} 12 & 7 - x \\ 24 & 16 \end{vmatrix}$$

$$= (-13 - x)((7 - x)(7 - x) - 4(16))$$
Theorem DMST [896]
$$+ (-8)(-1)(12(7 - x) - 4(24))$$

$$+ (-4)(12(16) - (7 - x)(24))$$

$$= 3 + 5x + x^{2} - x^{3}$$

 \triangle

 $= -(x-3)(x+1)^2$

 \boxtimes

 \square

The characteristic polynomial is our main computational tool for finding eigenvalues, and will sometimes be used to aid us in determining the properties of eigenvalues.

Theorem EMRCP Eigenvalues of a Matrix are Roots of Characteristic Polynomials

Suppose A is a square matrix. Then λ is an eigenvalue of A if and only if $p_A(\lambda) = 0$.

Proof Suppose A has size n.

 λ is an eigenvalue of A

$$\begin{array}{ll} \Longleftrightarrow & \text{there exists } \mathbf{x} \neq \mathbf{0} \text{ so that } A\mathbf{x} = \lambda \mathbf{x} & \text{Definition EEM [945]} \\ \Leftrightarrow & \text{there exists } \mathbf{x} \neq \mathbf{0} \text{ so that } A\mathbf{x} - \lambda \mathbf{x} = \mathbf{0} & \\ \Leftrightarrow & \text{there exists } \mathbf{x} \neq \mathbf{0} \text{ so that } A\mathbf{x} - \lambda I_n \mathbf{x} = \mathbf{0} & \\ \Leftrightarrow & \text{there exists } \mathbf{x} \neq \mathbf{0} \text{ so that } (A - \lambda I_n) \mathbf{x} = \mathbf{0} & \\ \Leftrightarrow & A - \lambda I_n \text{ is singular} & \\ \Leftrightarrow & \det (A - \lambda I_n) = \mathbf{0} & \\ \Leftrightarrow & p_A(\lambda) = \mathbf{0} & \\ \end{array}$$

eorem MMIM [475] eorem MMDAA [476] finition NM [173] eorem SMZD [935] finition CP [960]

Example EMS3 Eigenvalues of a matrix, size 3 In Example CPMS3 [961] we found the characteristic polynomial of

$$F = \begin{bmatrix} -13 & -8 & -4\\ 12 & 7 & 4\\ 24 & 16 & 7 \end{bmatrix}$$

to be $p_F(x) = -(x-3)(x+1)^2$. Factored, we can find all of its roots easily, they are x = 3 and x = -1. By Theorem EMRCP [962], $\lambda = 3$ and $\lambda = -1$ are both eigenvalues of F, and these are the only eigenvalues of F. We've found them all.

Let us now turn our attention to the computation of eigenvectors.

Definition EM Eigenspace of a Matrix

Suppose that A is a square matrix and λ is an eigenvalue of A. Then the **eigenspace** of A for λ , $\mathcal{E}_A(\lambda)$, is the set of all the eigenvectors of A for λ , together with the inclusion of the zero vector.

Example SEE [946] hinted that the set of eigenvectors for a single eigenvalue might have some closure properties, and with the addition of the non-eigenvector, $\mathbf{0}$, we indeed get a whole subspace.

Theorem EMS Eigenspace for a Matrix is a Subspace

Suppose A is a square matrix of size n and λ is an eigenvalue of A. Then the eigenspace

 $\mathcal{E}_{A}(\lambda)$ is a subspace of the vector space \mathbb{C}^{n} .

Proof We will check the three conditions of Theorem TSS [698]. First, Definition EM [963] explicitly includes the zero vector in $\mathcal{E}_A(\lambda)$, so the set is non-empty.

Suppose that $\mathbf{x}, \mathbf{y} \in \mathcal{E}_A(\lambda)$, that is, \mathbf{x} and \mathbf{y} are two eigenvectors of A for λ . Then

$A\left(\mathbf{x}+\mathbf{y}\right) = A\mathbf{x} + A\mathbf{y}$	Theorem MMDAA [476]
$=\lambda \mathbf{x}+\lambda \mathbf{y}$	\mathbf{x}, \mathbf{y} eigenvectors of A
$=\lambda\left(\mathbf{x}+\mathbf{y} ight)$	Property DVAC [206]

So either $\mathbf{x} + \mathbf{y} = \mathbf{0}$, or $\mathbf{x} + \mathbf{y}$ is an eigenvector of A for λ (Definition EEM [945]). So, in either event, $\mathbf{x} + \mathbf{y} \in \mathcal{E}_A(\lambda)$, and we have additive closure.

Suppose that $\alpha \in \mathbb{C}$, and that $\mathbf{x} \in \mathcal{E}_A(\lambda)$, that is, \mathbf{x} is an eigenvector of A for λ . Then

$A\left(\alpha\mathbf{x}\right) = \alpha\left(A\mathbf{x}\right)$	Theorem MMSMM [477]
$= \alpha \lambda \mathbf{x}$	\mathbf{x} an eigenvector of A
$=\lambda\left(lpha\mathbf{x} ight)$	Property SMAC [206]

So either $\alpha \mathbf{x} = \mathbf{0}$, or $\alpha \mathbf{x}$ is an eigenvector of A for λ (Definition EEM [945]). So, in either event, $\alpha \mathbf{x} \in \mathcal{E}_A(\lambda)$, and we have scalar closure.

With the three conditions of Theorem TSS [698] met, we know $\mathcal{E}_A(\lambda)$ is a subspace.

Theorem EMS [963] tells us that an eigenspace is a subspace (and hence a vector space in its own right). Our next theorem tells us how to quickly construct this subspace.

Theorem EMNS Eigenspace of a Matrix is a Null Space

Suppose A is a square matrix of size n and λ is an eigenvalue of A. Then

 $\mathcal{E}_A(\lambda) = \mathcal{N}(A - \lambda I_n)$

Proof The conclusion of this theorem is an equality of sets, so normally we would follow the advice of Definition SE [1604]. However, in this case we can construct a sequence of equivalences which will together provide the two subset inclusions we need. First, notice that $\mathbf{0} \in \mathcal{E}_A(\lambda)$ by Definition EM [963] and $\mathbf{0} \in \mathcal{N}(A - \lambda I_n)$ by Theorem HSC [147]. Now consider any nonzero vector $\mathbf{x} \in \mathbb{C}^n$,

$$\mathbf{x} \in \mathcal{E}_{A}(\lambda) \iff A\mathbf{x} = \lambda \mathbf{x} \qquad \text{Definition EM [963]}$$
$$\iff A\mathbf{x} - \lambda \mathbf{x} = \mathbf{0}$$
$$\iff A\mathbf{x} - \lambda I_{n}\mathbf{x} = \mathbf{0} \qquad \text{Theorem MMIM [475]}$$
$$\iff (A - \lambda I_{n})\mathbf{x} = \mathbf{0} \qquad \text{Theorem MMDAA [476]}$$
$$\iff \mathbf{x} \in \mathcal{N}(A - \lambda I_{n}) \qquad \text{Definition NSM [152]}$$

You might notice the close parallels (and differences) between the proofs of Theorem EMRCP [962] and Theorem EMNS [965]. Since Theorem EMNS [965] describes the set of

all the eigenvectors of A as a null space we can use techniques such as Theorem BNS [334] to provide concise descriptions of eigenspaces. Theorem EMNS [965] also provides a trivial proof for Theorem EMS [963].

Example ESMS3 Eigenspaces of a matrix, size 3

Example CPMS3 [961] and Example EMS3 [962] describe the characteristic polynomial and eigenvalues of the 3×3 matrix

$$F = \begin{bmatrix} -13 & -8 & -4\\ 12 & 7 & 4\\ 24 & 16 & 7 \end{bmatrix}$$

We will now take each eigenvalue in turn and compute its eigenspace. To do this, we row-reduce the matrix $F - \lambda I_3$ in order to determine solutions to the homogeneous system $\mathcal{LS}(F - \lambda I_3, \mathbf{0})$ and then express the eigenspace as the null space of $F - \lambda I_3$ (Theorem EMNS [965]). Theorem BNS [334] then tells us how to write the null space as the span of a basis.

$$\lambda = 3 \qquad F - 3I_3 = \begin{bmatrix} -16 & -8 & -4\\ 12 & 4 & 4\\ 24 & 16 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & \frac{1}{2}\\ 0 & 1 & -\frac{1}{2}\\ 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_F(3) = \mathcal{N}(F - 3I_3) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{2}\\ \frac{1}{2}\\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -1\\ 1\\ 2 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -1 \qquad F + 1I_3 = \begin{bmatrix} -12 & -8 & -4\\ 12 & 8 & 4\\ 24 & 16 & 8 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & \frac{2}{3} & \frac{1}{3}\\ 0 & 0 & 0\\ 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_F(-1) = \mathcal{N}(F + 1I_3) = \left\langle \left\{ \begin{bmatrix} -\frac{2}{3}\\ 1\\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{3}\\ 0\\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -2\\ 3\\ 0 \end{bmatrix}, \begin{bmatrix} -1\\ 0\\ 3 \end{bmatrix} \right\} \right\rangle$$

Eigenspaces in hand, we can easily compute eigenvectors by forming nontrivial linear combinations of the basis vectors describing each eigenspace. In particular, notice that we can "pretty up" our basis vectors by using scalar multiples to clear out fractions. More powerful scientific calculators, and most every mathematical software package, will compute eigenvalues of a matrix along with basis vectors of the eigenspaces. Be sure to understand how your device outputs complex numbers, since they are likely to occur. Also, the basis vectors will not necessarily look like the results of an application of Theorem BNS [334]. Duplicating the results of the next section (Subsection EE.ECEE [968]) with your device would be very good practice. See: Computation E.SAGE [1591]

Subsection ECEE Examples of Computing Eigenvalues and Eigenvectors

No theorems in this section, just a selection of examples meant to illustrate the range of possibilities for the eigenvalues and eigenvectors of a matrix. These examples can all be done by hand, though the computation of the characteristic polynomial would be very time-consuming and error-prone. It can also be difficult to factor an arbitrary polynomial, though if we were to suggest that most of our eigenvalues are going to be integers, then it can be easier to hunt for roots. These examples are meant to look similar to a concatenation of Example CPMS3 [961], Example EMS3 [962] and Example ESMS3 [966]. First, we will sneak in a pair of definitions so we can illustrate them throughout this sequence of examples.

Definition AME

Algebraic Multiplicity of an Eigenvalue

Suppose that A is a square matrix and λ is an eigenvalue of A. Then the **algebraic multiplicity** of λ , $\alpha_A(\lambda)$, is the highest power of $(x-\lambda)$ that divides the characteristic polynomial, $p_A(x)$.

(This definition contains Notation AME.)

Since an eigenvalue λ is a root of the characteristic polynomial, there is always a factor of $(x - \lambda)$, and the algebraic multiplicity is just the power of this factor in a factorization of $p_A(x)$. So in particular, $\alpha_A(\lambda) \geq 1$. Compare the definition of algebraic multiplicity with the next definition.

 \triangle

Definition GME Geometric Multiplicity of an Eigenvalue

Suppose that A is a square matrix and λ is an eigenvalue of A. Then the **geometric multiplicity** of λ , $\gamma_A(\lambda)$, is the dimension of the eigenspace $\mathcal{E}_A(\lambda)$. (This definition contains Notation GME.) \triangle

Since every eigenvalue must have at least one eigenvector, the associated eigenspace cannot be trivial, and so $\gamma_A(\lambda) \geq 1$.

Example EMMS4

Eigenvalue multiplicities, matrix of size 4

Consider the matrix

$$B = \begin{bmatrix} -2 & 1 & -2 & -4 \\ 12 & 1 & 4 & 9 \\ 6 & 5 & -2 & -4 \\ 3 & -4 & 5 & 10 \end{bmatrix}$$

then

$$p_B(x) = 8 - 20x + 18x^2 - 7x^3 + x^4 = (x - 1)(x - 2)^3$$

So the eigenvalues are $\lambda = 1, 2$ with algebraic multiplicities $\alpha_B(1) = 1$ and $\alpha_B(2) = 3$. Computing eigenvectors,

-

$$\lambda = 1 \qquad B - 1I_4 = \begin{bmatrix} -3 & 1 & -2 & -4 \\ 12 & 0 & 4 & 9 \\ 6 & 5 & -3 & -4 \\ 3 & -4 & 5 & 9 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

1

$$\mathcal{E}_{B}(1) = \mathcal{N}(B - 1I_{4}) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{3} \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -1 \\ 3 \\ 3 \\ 0 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 2 \qquad B - 2I_{4} = \begin{bmatrix} -4 & 1 & -2 & -4 \\ 12 & -1 & 4 & 9 \\ 6 & 5 & -4 & -4 \\ 3 & -4 & 5 & 8 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_{B}(2) = \mathcal{N}(B - 2I_{4}) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -1 \\ 2 \\ -1 \\ 2 \end{bmatrix} \right\} \right\rangle$$

So each eigenspace has dimension 1 and so $\gamma_B(1) = 1$ and $\gamma_B(2) = 1$. This example is of interest because of the discrepancy between the two multiplicities for $\lambda = 2$. In many of our examples the algebraic and geometric multiplicities will be equal for all of the eigenvalues (as it was for $\lambda = 1$ in this example), so keep this example in mind. We will have some explanations for this phenomenon later (see Example NDMS4 [1047]).

Example ESMS4 Eigenvalues, symmetric matrix of size 4 Consider the matrix

$$C = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

then

$$p_C(x) = -3 + 4x + 2x^2 - 4x^3 + x^4 = (x - 3)(x - 1)^2(x + 1)$$

So the eigenvalues are $\lambda = 3, 1, -1$ with algebraic multiplicities $\alpha_C(3) = 1, \alpha_C(1) = 2$ and $\alpha_C(-1) = 1$.

Computing eigenvectors,

$$\lambda = 3 \qquad C - 3I_4 = \begin{bmatrix} -2 & 0 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 0 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_C(3) = \mathcal{N}(C - 3I_4) = \left\langle \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 1 \qquad C - 1I_4 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_C(1) = \mathcal{N}(C - 1I_4) = \left\langle \left\{ \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = -1 \qquad C + 1I_4 = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_C(-1) = \mathcal{N}(C + 1I_4) = \left\langle \left\{ \left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \right\} \right\} \right\rangle$$

So the eigenspace dimensions yield geometric multiplicities $\gamma_C(3) = 1$, $\gamma_C(1) = 2$ and $\gamma_C(-1) = 1$, the same as for the algebraic multiplicities. This example is of interest because A is a symmetric matrix, and will be the subject of Theorem HMRE [1020].

Example HMEM5

High multiplicity eigenvalues, matrix of size 5

Consider the matrix

$$E = \begin{bmatrix} 29 & 14 & 2 & 6 & -9 \\ -47 & -22 & -1 & -11 & 13 \\ 19 & 10 & 5 & 4 & -8 \\ -19 & -10 & -3 & -2 & 8 \\ 7 & 4 & 3 & 1 & -3 \end{bmatrix}$$

then

$$p_E(x) = -16 + 16x + 8x^2 - 16x^3 + 7x^4 - x^5 = -(x-2)^4(x+1)$$

So the eigenvalues are $\lambda = 2$, -1 with algebraic multiplicities $\alpha_E(2) = 4$ and $\alpha_E(-1) = 1$. Computing eigenvectors,

_

$$\lambda = -1 \qquad E + 1I_5 = \begin{bmatrix} 30 & 14 & 2 & 6 & -9 \\ -47 & -21 & -1 & -11 & 13 \\ 19 & 10 & 6 & 4 & -8 \\ -19 & -10 & -3 & -1 & 8 \\ 7 & 4 & 3 & 1 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & -4 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_E(-1) = \mathcal{N}(E + 1I_5) = \left\langle \left\{ \begin{bmatrix} -2 \\ 4 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle$$

So the eigenspace dimensions yield geometric multiplicities $\gamma_E(2) = 2$ and $\gamma_E(-1) = 1$. This example is of interest because $\lambda = 2$ has such a large algebraic multiplicity, which is also not equal to its geometric multiplicity.

Example CEMS6 Complex eigenvalues, matrix of size 6 Consider the matrix

$$F = \begin{bmatrix} -59 & -34 & 41 & 12 & 25 & 30 \\ 1 & 7 & -46 & -36 & -11 & -29 \\ -233 & -119 & 58 & -35 & 75 & 54 \\ 157 & 81 & -43 & 21 & -51 & -39 \\ -91 & -48 & 32 & -5 & 32 & 26 \\ 209 & 107 & -55 & 28 & -69 & -50 \end{bmatrix}$$

then

$$p_F(x) = -50 + 55x + 13x^2 - 50x^3 + 32x^4 - 9x^5 + x^6$$

= $(x - 2)(x + 1)(x^2 - 4x + 5)^2$
= $(x - 2)(x + 1)((x - (2 + i))(x - (2 - i)))^2$
= $(x - 2)(x + 1)(x - (2 + i))^2(x - (2 - i))^2$

So the eigenvalues are $\lambda = 2, -1, 2 + i, 2 - i$ with algebraic multiplicities $\alpha_F(2) = 1, \alpha_F(-1) = 1, \alpha_F(2+i) = 2$ and $\alpha_F(2-i) = 2$.

Computing eigenvectors,

$$\lambda = 2$$

$$F - 2I_{6} = \begin{bmatrix} -61 & -34 & 41 & 12 & 25 & 30 \\ 1 & 5 & -46 & -36 & -11 & -29 \\ -233 & -119 & 56 & -35 & 75 & 54 \\ 157 & 81 & -43 & 19 & -51 & -39 \\ -91 & -48 & 32 & -5 & 30 & 26 \\ 209 & 107 & -55 & 28 & -69 & -52 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & 1 & 0 & -\frac{1}{5} \\ 0 & 0 & 0 & 0 & 1 & \frac{4}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_{F}(2) = \mathcal{N}(F - 2I_{6}) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{5} \\ 0 \\ -\frac{3}{15} \\ \frac{1}{5} \\ -\frac{4}{5} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -1 \\ 0 \\ -3 \\ 1 \\ -4 \\ 5 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -1$$

$$F + 1I_6 = \begin{bmatrix} -58 & -34 & 41 & 12 & 25 & 30 \\ 1 & 8 & -46 & -36 & -11 & -29 \\ -233 & -119 & 59 & -35 & 75 & 54 \\ 157 & 81 & -43 & 22 & -51 & -39 \\ -91 & -48 & 32 & -5 & 33 & 26 \\ 209 & 107 & -55 & 28 & -69 & -49 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathcal{E}_{F}(-1) = \mathcal{N}(F + I_{6}) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{2} \\ \frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -1 \\ 3 \\ -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 2 + i$$

$$F - (2+i)I_6 = \begin{bmatrix} -61 - i & -34 & 41 & 12 & 25 & 30 \\ 1 & 5 - i & -46 & -36 & -11 & -29 \\ -233 & -119 & 56 - i & -35 & 75 & 54 \\ 157 & 81 & -43 & 19 - i & -51 & -39 \\ -91 & -48 & 32 & -5 & 30 - i & 26 \\ 209 & 107 & -55 & 28 & -69 & -52 - i \end{bmatrix}$$

$$\xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \frac{1}{5}(7+i) \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{5}(-9-2i) \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_{F}(2+i) = \mathcal{N}(F - (2+i)I_{6}) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{5}(7+i) \\ \frac{1}{5}(9+2i) \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -7-i \\ 9+2i \\ -5 \\ 5 \\ -5 \\ 5 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 2 - i$$

$$F - (2 - i)I_6 = \begin{bmatrix} -61 + i & -34 & 41 & 12 & 25 & 30 \\ 1 & 5 + i & -46 & -36 & -11 & -29 \\ -233 & -119 & 56 + i & -35 & 75 & 54 \\ 157 & 81 & -43 & 19 + i & -51 & -39 \\ -91 & -48 & 32 & -5 & 30 + i & 26 \\ 209 & 107 & -55 & 28 & -69 & -52 + i \end{bmatrix}$$

$$\xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \frac{1}{5}(7 - i) \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{5}(-9 + 2i) \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_F(2-i) = \mathcal{N}(F-(2-i)I_6) = \left\langle \left\{ \begin{bmatrix} \frac{1}{5}(-7+i)\\ \frac{1}{5}(9-2i)\\ -1\\ 1\\ -1\\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -7+i\\ 9-2i\\ -5\\ 5\\ -5\\ 5 \end{bmatrix} \right\} \right\rangle$$

So the eigenspace dimensions yield geometric multiplicities $\gamma_F(2) = 1$, $\gamma_F(-1) = 1$, $\gamma_F(2+i) = 1$ and $\gamma_F(2-i) = 1$. This example demonstrates some of the possibilities for the appearance of complex eigenvalues, even when all the entries of the matrix are real. Notice how all the numbers in the analysis of $\lambda = 2 - i$ are conjugates of the corresponding number in the analysis of $\lambda = 2 + i$. This is the content of the upcoming Theorem ERMCP [1012].

Example DEMS5 Distinct eigenvalues, matrix of size 5

Consider the matrix

$$H = \begin{bmatrix} 15 & 18 & -8 & 6 & -5\\ 5 & 3 & 1 & -1 & -3\\ 0 & -4 & 5 & -4 & -2\\ -43 & -46 & 17 & -14 & 15\\ 26 & 30 & -12 & 8 & -10 \end{bmatrix}$$

then

$$p_H(x) = -6x + x^2 + 7x^3 - x^4 - x^5 = x(x-2)(x-1)(x+1)(x+3)$$

So the eigenvalues are $\lambda = 2, 1, 0, -1, -3$ with algebraic multiplicities $\alpha_H(2) = 1, \alpha_H(1) = 1, \alpha_H(0) = 1, \alpha_H(-1) = 1$ and $\alpha_H(-3) = 1$.

Computing eigenvectors,

$$\lambda = 2 \qquad H - 2I_5 = \begin{bmatrix} 13 & 18 & -8 & 6 & -5 \\ 5 & 1 & 1 & -1 & -3 \\ 0 & -4 & 3 & -4 & -2 \\ -43 & -46 & 17 & -16 & 15 \\ 26 & 30 & -12 & 8 & -12 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_H(2) = \mathcal{N}(H - 2I_5) = \left\langle \left\{ \begin{bmatrix} 1 \\ -1 \\ -2 \\ -1 \\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 1 \qquad H - 1I_5 = \begin{bmatrix} 14 & 18 & -8 & 6 & -5\\ 5 & 2 & 1 & -1 & -3\\ 0 & -4 & 4 & -4 & -2\\ -43 & -46 & 17 & -15 & 15\\ 26 & 30 & -12 & 8 & -11 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{2}\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & \frac{1}{2}\\ 0 & 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_{H}(1) = \mathcal{N}(H - 1I_{5}) = \left\langle \left\{ \begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ -1 \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ -2 \\ 2 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 0 \qquad H - 0I_5 = \begin{bmatrix} 15 & 18 & -8 & 6 & -5 \\ 5 & 3 & 1 & -1 & -3 \\ 0 & -4 & 5 & -4 & -2 \\ -43 & -46 & 17 & -14 & 15 \\ 26 & 30 & -12 & 8 & -10 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_H(0) = \mathcal{N}(H - 0I_5) = \left\langle \left\{ \begin{bmatrix} -1 \\ 2 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -1 \quad H + 1I_5 = \begin{bmatrix} 16 & 18 & -8 & 6 & -5 \\ 5 & 4 & 1 & -1 & -3 \\ 0 & -4 & 6 & -4 & -2 \\ -43 & -46 & 17 & -13 & 15 \\ 26 & 30 & -12 & 8 & -9 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & -1/2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_{H}(-1) = \mathcal{N}(H+1I_{5}) = \left\langle \left\{ \begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \\ -\frac{1}{2} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 2 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -3 \qquad H + 3I_5 = \begin{bmatrix} 18 & 18 & -8 & 6 & -5\\ 5 & 6 & 1 & -1 & -3\\ 0 & -4 & 8 & -4 & -2\\ -43 & -46 & 17 & -11 & 15\\ 26 & 30 & -12 & 8 & -7 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & -1\\ 0 & 1 & 0 & 0 & \frac{1}{2}\\ 0 & 0 & 1 & 0 & 1\\ 0 & 0 & 0 & 1 & 2\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_H (-3) = \mathcal{N}(H + 3I_5) = \left\langle \left\{ \begin{bmatrix} 1\\ -\frac{1}{2}\\ -1\\ -2\\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -2\\ 1\\ 2\\ 4\\ -2 \end{bmatrix} \right\} \right\rangle$$

So the eigenspace dimensions yield geometric multiplicities $\gamma_H(2) = 1$, $\gamma_H(1) = 1$, $\gamma_H(0) = 1$, $\gamma_H(-1) = 1$ and $\gamma_H(-3) = 1$, identical to the algebraic multiplicities. This example is of interest for two reasons. First, $\lambda = 0$ is an eigenvalue, illustrating the upcoming Theorem SMZE [1003]. Second, all the eigenvalues are distinct, yielding algebraic and geometric multiplicities of 1 for each eigenvalue, illustrating Theorem DED [1048].

Subsection READ Reading Questions

Suppose A is the 2×2 matrix

$$A = \begin{bmatrix} -5 & 8\\ -4 & 7 \end{bmatrix}$$

- 1. Find the eigenvalues of A.
- 2. Find the eigenspaces of A.
- 3. For the polynomial $p(x) = 3x^2 x + 2$, compute p(A).

Subsection EXC Exercises

C10 Find the characteristic polynomial of the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Contributed by Chris Black Solution [989] C11 Find the characteristic polynomial of the matrix $A = \begin{bmatrix} 3 & 2 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$.

Contributed by Chris Black Solution [989]

C12 Find the characteristic polynomial of the matrix $A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 \\ 3 & 1 & 0 & 1 \end{bmatrix}$.

Contributed by Chris Black Solution [989]

C19 Find the eigenvalues, eigenspaces, algebraic multiplicities and geometric multiplicities for the matrix below. It is possible to do all these computations by hand, and it would be instructive to do so.

$$C = \begin{bmatrix} -1 & 2\\ -6 & 6 \end{bmatrix}$$

Contributed by Robert Beezer Solution [989]

C20 Find the eigenvalues, eigenspaces, algebraic multiplicities and geometric multiplicities for the matrix below. It is possible to do all these computations by hand, and it would be instructive to do so.

$$B = \begin{bmatrix} -12 & 30\\ -5 & 13 \end{bmatrix}$$

Contributed by Robert Beezer Solution [990]

C21 The matrix A below has $\lambda = 2$ as an eigenvalue. Find the geometric multiplicity of $\lambda = 2$ using your calculator only for row-reducing matrices.

$$A = \begin{bmatrix} 18 & -15 & 33 & -15 \\ -4 & 8 & -6 & 6 \\ -9 & 9 & -16 & 9 \\ 5 & -6 & 9 & -4 \end{bmatrix}$$

Contributed by Robert Beezer Solution [992]

C22 Without using a calculator, find the eigenvalues of the matrix B.

$$B = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [992]

C23 Find the eigenvalues, eigenspaces, algebraic and geometric multiplicities for $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Contributed by Chris Black Solution [993]

C24 Find the eigenvalues, eigenspaces, algebraic and geometric multiplicities for $A = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$. Contributed by Chris Black Solution [993]

C25 Find the eigenvalues, eigenspaces, algebraic and geometric multiplicities for the 3×3 identity matrix I_3 . Do your results make sense? Contributed by Chris Black Solution [993]

C26 For matrix $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, the characteristic polynomial of A is $p_A(\lambda) = (4-x)(1-x)^2$. Find the eigenvalues and corresponding eigenspaces of A. Contributed by Chris Black Solution [994]

C27 For matrix
$$A = \begin{bmatrix} 0 & 4 & -1 & 1 \\ -2 & 6 & -1 & 1 \\ -2 & 8 & -1 & -1 \\ -2 & 8 & -3 & 1 \end{bmatrix}$$
, the characteristic polynomial of A is
$$p_A(\lambda) = (x+2)(x-2)^2(x-4).$$
Find the eigenvalues and corresponding eigenspaces of A .
Contributed by Chris Black Solution [994]

M60 Repeat Example CAEHW [955] by choosing $\mathbf{x} = \begin{bmatrix} 0 \\ 8 \\ 2 \\ 1 \\ 2 \end{bmatrix}$ and then arrive at an eigenvalue and eigenvector of the matrix A. The hard way.

Contributed by Robert Beezer Solution [995]

T10 A matrix A is idempotent if $A^2 = A$. Show that the only possible eigenvalues of an idempotent matrix are $\lambda = 0$ and $\lambda = 1$. Then give an example of a matrix that is idempotent and has both of these two values as eigenvalues. Contributed by Robert Beezer Solution [997]

T15 The characteristic polynomial of the square matrix A is usually defined as $r_A(x) = \det(xI_n - A)$. Find a specific relationship between our characteristic polynomial, $p_A(x)$,

and $r_A(x)$, give a proof of your relationship, and use this to explain why Theorem EMRCP [962] can remain essentially unchanged with either definition. Explain the advantages of each definition over the other. (Computing with both definitions, for a 2 × 2 and a 3 × 3 matrix, might be a good way to start.)

Contributed by Robert Beezer Solution [998]

T20 Suppose that λ and ρ are two different eigenvalues of the square matrix A. Prove that the intersection of the eigenspaces for these two eigenvalues is trivial. That is, $\mathcal{E}_A(\lambda) \cap \mathcal{E}_A(\rho) = \{\mathbf{0}\}.$

Contributed by Robert Beezer Solution [999]

Subsection SOL Solutions

C10 Contributed by Chris Black Statement [984] Answer: $p_A(x) = -2 - 5x + x^2$

C11 Contributed by Chris Black Statement [984] Answer: $p_A(x) = -5 + 4x^2 - x^3$.

C12 Contributed by Chris Black Statement [984] Answer: $p_A(x) = 2 + 2x - 2x^2 - 3x^3 + x^4$.

C19 Contributed by Robert Beezer Statement [984] First compute the characteristic polynomial,

$$p_{C}(x) = \det (C - xI_{2})$$

Definition CP [960]
$$= \begin{vmatrix} -1 - x & 2 \\ -6 & 6 - x \end{vmatrix}$$

$$= (-1 - x)(6 - x) - (2)(-6)$$

$$= x^{2} - 5x + 6$$

$$= (x - 3)(x - 2)$$

So the eigenvalues of C are the solutions to $p_C(x) = 0$, namely, $\lambda = 2$ and $\lambda = 3$.

To obtain the eigenspaces, construct the appropriate singular matrices and find expressions for the null spaces of these matrices.

$$\lambda = 2$$

$$C - (2)I_2 = \begin{bmatrix} -3 & 2 \\ -6 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -\frac{2}{3} \\ 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_C(2) = \mathcal{N}(C - (2)I_2) = \left\langle \left\{ \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 3$$

$$C - (3)I_2 = \begin{bmatrix} -4 & 2\\ -6 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_C(3) = \mathcal{N}(C - (3)I_2) = \left\langle \left\{ \begin{bmatrix} 1\\2\\1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 1\\2\\2 \end{bmatrix} \right\} \right\rangle$$

C20Contributed by Robert BeezerStatement [985]The characteristic polynomial of B is

$$p_B(x) = \det (B - xI_2) \\ = \begin{vmatrix} -12 - x & 30 \\ -5 & 13 - x \end{vmatrix}$$

Definition CP [960]

$$= (-12 - x)(13 - x) - (30)(-5)$$

Theorem DMST [896]
$$= x^{2} - x - 6$$

$$= (x - 3)(x + 2)$$

From this we find eigenvalues $\lambda = 3, -2$ with algebraic multiplicities $\alpha_B(3) = 1$ and $\alpha_B(-2) = 1$.

For eigenvectors and geometric multiplicities, we study the null spaces of $B - \lambda I_2$ (Theorem EMNS [965]).

$$\lambda = 3 \qquad B - 3I_2 = \begin{bmatrix} -15 & 30 \\ -5 & 10 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_B(3) = \mathcal{N}(B - 3I_2) = \left\langle \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -2 \qquad B + 2I_2 = \begin{bmatrix} -10 & 30 \\ -5 & 15 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_B(-2) = \mathcal{N}(B + 2I_2) = \left\langle \left\{ \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\} \right\rangle$$

Each eigenspace has dimension one, so we have geometric multiplicities $\gamma_B(3) = 1$ and $\gamma_B(-2) = 1$.

C21 Contributed by Robert Beezer Statement [985]

If $\lambda = 2$ is an eigenvalue of A, the matrix $A - 2I_4$ will be singular, and its null space will be the eigenspace of A. So we form this matrix and row-reduce,

$$A - 2I_4 = \begin{bmatrix} 16 & -15 & 33 & -15 \\ -4 & 6 & -6 & 6 \\ -9 & 9 & -18 & 9 \\ 5 & -6 & 9 & -6 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

With two free variables, we know a basis of the null space (Theorem BNS [334]) will contain two vectors. Thus the null space of $A - 2I_4$ has dimension two, and so the eigenspace of $\lambda = 2$ has dimension two also (Theorem EMNS [965]), $\gamma_A(2) = 2$.

C22 Contributed by Robert Beezer Statement [985] The characteristic polynomial (Definition CP [960]) is

$$p_B(x) = \det (B - xI_2)$$

= $\begin{vmatrix} 2 - x & -1 \\ 1 & 1 - x \end{vmatrix}$
= $(2 - x)(1 - x) - (1)(-1)$
= $x^2 - 3x + 3$
= $\left(x - \frac{3 + \sqrt{3}i}{2}\right) \left(x - \frac{3 - \sqrt{3}i}{2}\right)$

Theorem DMST [896]

where the factorization can be obtained by finding the roots of $p_B(x) = 0$ with the quadratic equation. By Theorem EMRCP [962] the eigenvalues of B are the complex numbers $\lambda_1 = \frac{3+\sqrt{3}i}{2}$ and $\lambda_2 = \frac{3-\sqrt{3}i}{2}$.

C23 Contributed by Chris Black Statement [986] Eigenvalue: $\lambda = 0$ $\mathcal{E}_A(0) = \left\langle \begin{bmatrix} -1\\1 \end{bmatrix} \right\rangle, \, \alpha_A(0) = 1, \, \gamma_A(0) = 1$ Eigenvalue: $\lambda = 2$ $\mathcal{E}_A(2) = \left\langle \begin{bmatrix} 1\\1 \end{bmatrix} \right\rangle, \, \alpha_A(2) = 1, \, \gamma_A(2) = 1$

C24 Contributed by Chris Black Statement [986] Eigenvalue: $\lambda = 0$

$$\mathcal{E}_{A}(0) = \left\langle \begin{bmatrix} 1\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} -1\\0\\1\\1 \end{bmatrix} \right\rangle, \alpha_{A}(0) = 2, \gamma_{A}(0) = 2$$

Eigenvalue: $\lambda = 3$
$$\mathcal{E}_{A}(3) = \left\langle \begin{bmatrix} 1\\-1\\1\\1 \end{bmatrix} \right\rangle, \alpha_{A}(3) = 1, \gamma_{A}(3) = 1$$

C25 Contributed by Chris Black Statement [986] The characteristic polynomial for $A = I_3$ is $p_{I_3}(x) = (1-x)^3$, which has eigenvalue $\lambda = 1$ with algebraic multiplicity $\alpha_A(1) = 3$. Looking for eigenvectors, we find that $A - \lambda I =$

 $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. The nullspace of this matrix is all of \mathbb{C}^3 , so that the eigenspace is $\mathcal{E}_{I_3}(1) =$

 $\left\langle \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\rangle$, and the geometric multiplicity is $\gamma_A(1) = 3$.

Does this make sense? Yes! Every vector \mathbf{x} is a solution to $I_3\mathbf{x} = 1\mathbf{x}$, so every nonzero vector is an eigenvector with eigenvalue 1. Since every vector is unchanged when multiplied by I_3 , it makes sense that $\lambda = 1$ is the only eigenvalue.

C26 Contributed by Chris Black Statement [986] Since we are given that the characteristic polynomial of A is $p_A(x) = (4 - x)(1 - x)^2$, we see that the eigenvalues are $\lambda = 4$ with algebraic multiplicity $\alpha_A(4) = 1$ and $\lambda = 1$ with algebraic multiplicity $\alpha_A(1) = 2$. The corresponding eigenspaces are

$$\mathcal{E}_{A}(4) = \left\langle \begin{bmatrix} 1\\1\\1 \end{bmatrix} \right\rangle \qquad \qquad \mathcal{E}_{A}(1) = \left\langle \begin{bmatrix} 1\\-1\\0 \end{bmatrix}, \begin{bmatrix} 1\\0\\-1 \end{bmatrix} \right\rangle$$

C27Contributed by Chris Black Statement [987] Since we are given that the characteristic polynomial of A is $p_A(x) = (x+2)(x-2)^2(x-4)$, we see that the eigenvalues are $\lambda = -2$, $\lambda = 2$ and $\lambda = 4$. The eigenspaces are

$$\mathcal{E}_{A}(-2) = \left\langle \begin{bmatrix} 0\\0\\1\\1 \end{bmatrix} \right\rangle$$
$$\mathcal{E}_{A}(2) = \left\langle \begin{bmatrix} 1\\1\\2\\0 \end{bmatrix}, \begin{bmatrix} 3\\1\\0\\2 \end{bmatrix} \right\rangle$$
$$\mathcal{E}_{A}(4) = \left\langle \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \right\rangle$$

M60 Contributed by Robert Beezer Statement [987] Form the matrix C whose columns are \mathbf{x} , $A\mathbf{x}$, $A^2\mathbf{x}$, $A^3\mathbf{x}$, $A^4\mathbf{x}$, $A^5\mathbf{x}$ and row-reduce the matrix,

The simplest possible relation of linear dependence on the columns of C comes from using scalars $\alpha_4 = 1$ and $\alpha_5 = \alpha_6 = 0$ for the free variables in a solution to $\mathcal{LS}(C, \mathbf{0})$. The remainder of this solution is $\alpha_1 = 3$, $\alpha_2 = -1$, $\alpha_3 = -3$. This solution gives rise to the polynomial

$$p(x) = 3 - x - 3x^{2} + x^{3} = (x - 3)(x - 1)(x + 1)$$

which then has the property that $p(A)\mathbf{x} = \mathbf{0}$.

No matter how you choose to order the factors of p(x), the value of k (in the language of Theorem EMHE [953] and Example CAEHW [955]) is k = 2. For each of the three possibilities, we list the resulting eigenvector and the associated eigenvalue:

$$(C - 3I_5)(C - I_5)\mathbf{z} = \begin{bmatrix} 8\\ 8\\ -24\\ 8 \end{bmatrix} \qquad \lambda = -1$$
$$(C - 3I_5)(C + I_5)\mathbf{z} = \begin{bmatrix} 20\\ -20\\ 20\\ -40\\ 20 \end{bmatrix} \qquad \lambda = 1$$

$$(C+I_5)(C-I_5)\mathbf{z} = \begin{bmatrix} 32\\16\\48\\-48\\48 \end{bmatrix} \qquad \lambda = 3$$

Note that each of these eigenvectors can be simplified by an appropriate scalar multiple, but we have shown here the actual vector obtained by the product specified in the theorem.

T10 Contributed by Robert Beezer Statement [987] Suppose that λ is an eigenvalue of A. Then there is an eigenvector \mathbf{x} , such that $A\mathbf{x} = \lambda \mathbf{x}$. We have,

$\lambda \mathbf{x} = A\mathbf{x}$	\mathbf{x} eigenvector of A
$= A^2 \mathbf{x}$	A is idempotent
$=A(A\mathbf{x})$	
$=A(\lambda \mathbf{x})$	\mathbf{x} eigenvector of A
$=\lambda(A\mathbf{x})$	Theorem MMSMM [477]
$=\lambda(\lambda \mathbf{x})$	\mathbf{x} eigenvector of A
$=\lambda^2 \mathbf{x}$	

From this we get

$$\mathbf{0} = \lambda^2 \mathbf{x} - \lambda \mathbf{x}$$

 $= (\lambda^2 - \lambda)\mathbf{x}$ Property DSAC [206]

Since **x** is an eigenvector, it is nonzero, and Theorem SMEZV [679] leaves us with the conclusion that $\lambda^2 - \lambda = 0$, and the solutions to this quadratic polynomial equation in λ are $\lambda = 0$ and $\lambda = 1$.

The matrix



is idempotent (check this!) and since it is a diagonal matrix, its eigenvalues are the diagonal entries, $\lambda = 0$ and $\lambda = 1$, so each of these possible values for an eigenvalue of an idempotent matrix actually occurs as an eigenvalue of some idempotent matrix. So we cannot state any stronger conclusion about the eigenvalues of an idempotent matrix, and we can say that this theorem is the "best possible."

T15 Contributed by Robert Beezer Statement [987]

Note in the following that the scalar multiple of a matrix is equivalent to multiplying each of the rows by that scalar, so we actually apply Theorem DRCM [922] multiple times below (and are passing up an opportunity to do a proof by induction in the process, which maybe you'd like to do yourself?).

$$p_A(x) = \det (A - xI_n)$$
$$= \det ((-1)(xI_n - A))$$

Definition CP [960] Definition MSM [428]

$$= (-1)^{n} \det (xI_{n} - A)$$
 Theorem DRCM [922]
$$= (-1)^{n} r_{A}(x)$$

Since the polynomials are scalar multiples of each other, their roots will be identical, so either polynomial could be used in Theorem EMRCP [962].

Computing by hand, our definition of the characteristic polynomial is easier to use, as you only need to subtract x down the diagonal of the matrix before computing the determinant. However, the price to be paid is that for odd values of n, the coefficient of x^n is -1, while $r_A(x)$ always has the coefficient 1 for x^n (we say $r_A(x)$ is "monic.")

T20 Contributed by Robert Beezer Statement [988] This problem asks you to prove that two sets are equal, so use Definition SE [1604].

First show that $\{\mathbf{0}\} \subseteq \mathcal{E}_A(\lambda) \cap \mathcal{E}_A(\rho)$. Choose $\mathbf{x} \in \{\mathbf{0}\}$. Then $\mathbf{x} = \mathbf{0}$. Eigenspaces are subspaces (Theorem EMS [963]), so both $\mathcal{E}_A(\lambda)$ and $\mathcal{E}_A(\rho)$ contain the zero vector, and therefore $\mathbf{x} \in \mathcal{E}_A(\lambda) \cap \mathcal{E}_A(\rho)$ (Definition SI [1608]).

To show that $\mathcal{E}_A(\lambda) \cap \mathcal{E}_A(\rho) \subseteq \{\mathbf{0}\}$, suppose that $\mathbf{x} \in \mathcal{E}_A(\lambda) \cap \mathcal{E}_A(\rho)$. Then \mathbf{x} is an eigenvector of A for both λ and ρ (Definition SI [1608]) and so

$$\mathbf{x} = 1\mathbf{x} \qquad \text{Property O [661]} \\ = \frac{1}{\lambda - \rho} (\lambda - \rho) \mathbf{x} \qquad \lambda \neq \rho, \ \lambda - \rho \neq 0 \\ = \frac{1}{\lambda - \rho} (\lambda \mathbf{x} - \rho \mathbf{x}) \qquad \text{Property DSAC [206]}$$

$$= \frac{1}{\lambda - \rho} (A\mathbf{x} - A\mathbf{x})$$
$$= \frac{1}{\lambda - \rho} (\mathbf{0})$$
$$= \mathbf{0}$$

 ${\bf x}$ eigenvector of A for $\lambda,\,\rho$

Theorem ZVSM [677]

So $\mathbf{x} = \mathbf{0}$, and trivially, $\mathbf{x} \in \{\mathbf{0}\}$.

Section PEE Properties of Eigenvalues and Eigenvectors

The previous section introduced eigenvalues and eigenvectors, and concentrated on their existence and determination. This section will be more about theorems, and the various properties eigenvalues and eigenvectors enjoy. Like a good 4×100 meter relay, we will lead-off with one of our better theorems and save the very best for the anchor leg.

Theorem EDELI

Eigenvectors with Distinct Eigenvalues are Linearly Independent

Suppose that A is an $n \times n$ square matrix and $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_p\}$ is a set of eigenvectors with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ such that $\lambda_i \neq \lambda_j$ whenever $i \neq j$. Then S is a linearly independent set.

Proof If p = 1, then the set $S = {\mathbf{x}_1}$ is linearly independent since eigenvectors are nonzero (Definition EEM [945]), so assume for the remainder that $p \ge 2$.

We will prove this result by contradiction (Technique CD [1623]). Suppose to the contrary that S is a linearly dependent set. Define $S_i = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_i\}$ and let k be an integer such that $S_{k-1} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{k-1}\}$ is linearly independent and $S_k = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_k\}$ is linearly dependent. We have to ask if there is even such an integer k? First, since eigenvectors are nonzero, the set $\{\mathbf{x}_1\}$ is linearly independent. Since we are assuming that $S = S_p$ is linearly dependent, there must be an integer $k, 2 \leq k \leq p$, where

the sets S_i transition from linear independence to linear dependence (and stay that way). In other words, \mathbf{x}_k is the vector with the smallest index that is a linear combination of just vectors with smaller indices.

Since $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_k\}$ is linearly dependent there are scalars, $a_1, a_2, a_3, \dots, a_k$, some non-zero (Definition LI [735]), so that

$$\mathbf{0} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 + \dots + a_k \mathbf{x}_k$$

Then,

$$\begin{aligned} \mathbf{0} &= (A - \lambda_k I_n) \, \mathbf{0} & \text{Theorem} \\ &= (A - \lambda_k I_n) \left(a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 + \dots + a_k \mathbf{x}_k \right) & \text{Defin} \\ &= (A - \lambda_k I_n) a_1 \mathbf{x}_1 + (A - \lambda_k I_n) a_2 \mathbf{x}_2 + \dots + (A - \lambda_k I_n) a_k \mathbf{x}_k & \text{Theorem} \\ &= a_1 \left(A - \lambda_k I_n \right) \mathbf{x}_1 + a_2 \left(A - \lambda_k I_n \right) \mathbf{x}_2 + \dots + a_k \left(A - \lambda_k I_n \right) \mathbf{x}_k & \text{Theorem} \\ &= a_1 \left(A \mathbf{x}_1 - \lambda_k I_n \mathbf{x}_1 \right) + a_2 \left(A \mathbf{x}_2 - \lambda_k I_n \mathbf{x}_2 \right) + \dots + a_k \left(A \mathbf{x}_k - \lambda_k I_n \mathbf{x}_k \right) & \text{Theorem} \\ &= a_1 \left(A \mathbf{x}_1 - \lambda_k \mathbf{x}_1 \right) + a_2 \left(A \mathbf{x}_2 - \lambda_k \mathbf{x}_2 \right) + \dots + a_k \left(A \mathbf{x}_k - \lambda_k \mathbf{x}_k \right) & \text{Theorem} \\ &= a_1 \left(\lambda_1 \mathbf{x}_1 - \lambda_k \mathbf{x}_1 \right) + a_2 \left(\lambda_2 \mathbf{x}_2 - \lambda_k \mathbf{x}_2 \right) + \dots + a_k \left(\lambda_k \mathbf{x}_k - \lambda_k \mathbf{x}_k \right) & \text{Defin} \\ &= a_1 \left(\lambda_1 - \lambda_k \right) \mathbf{x}_1 + a_2 \left(\lambda_2 - \lambda_k \right) \mathbf{x}_2 + \dots + a_k \left(\lambda_k - \lambda_k \right) \mathbf{x}_k & \text{Theorem} \\ &= a_1 \left(\lambda_1 - \lambda_k \right) \mathbf{x}_1 + a_2 \left(\lambda_2 - \lambda_k \right) \mathbf{x}_2 + \dots + a_k \left(0 \right) \mathbf{x}_k & \text{Theorem} \\ &= a_1 \left(\lambda_1 - \lambda_k \right) \mathbf{x}_1 + a_2 \left(\lambda_2 - \lambda_k \right) \mathbf{x}_2 + \dots + a_k \left(\lambda_{k-1} - \lambda_k \right) \mathbf{x}_{k-1} + \mathbf{0} & \text{Theorem} \\ &= a_1 \left(\lambda_1 - \lambda_k \right) \mathbf{x}_1 + a_2 \left(\lambda_2 - \lambda_k \right) \mathbf{x}_2 + \dots + a_{k-1} \left(\lambda_{k-1} - \lambda_k \right) \mathbf{x}_{k-1} & \text{Prop} \end{aligned}$$

Theorem ZVSM [677] Definition RLD [735] Theorem MMDAA [476] Theorem MMDAA [476] Theorem MMDAA [476] Theorem MMIM [475] Definition EEM [945] Theorem MMDAA [476] Property AICN [1597] Theorem ZSSM [676] Property Z [660] This is a relation of linear dependence on the linearly independent set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{k-1}\}$, so the scalars must all be zero. That is, $a_i (\lambda_i - \lambda_k) = 0$ for $1 \le i \le k - 1$. However, we have the hypothesis that the eigenvalues are distinct, so $\lambda_i \ne \lambda_k$ for $1 \le i \le k - 1$. Thus $a_i = 0$ for $1 \le i \le k - 1$.

This reduces the original relation of linear dependence on $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_k\}$ to the simpler equation $a_k \mathbf{x}_k = \mathbf{0}$. By Theorem SMEZV [679] we conclude that $a_k = 0$ or $\mathbf{x}_k = \mathbf{0}$. Eigenvectors are never the zero vector (Definition EEM [945]), so $a_k = 0$. So all of the scalars $a_i, 1 \leq i \leq k$ are zero, contradicting their introduction as the scalars creating a nontrivial relation of linear dependence on the set $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_k\}$. With a contradiction in hand, we conclude that S must be linearly independent.

There is a simple connection between the eigenvalues of a matrix and whether or not the matrix is nonsingular.

Theorem SMZE

Singular Matrices have Zero Eigenvalues

Suppose A is a square matrix. Then A is singular if and only if $\lambda = 0$ is an eigenvalue of A.

Proof We have the following equivalences:

A is singular
$$\iff$$
 there exists $\mathbf{x} \neq \mathbf{0}$, $A\mathbf{x} = \mathbf{0}$ Definition NSM [152] \iff there exists $\mathbf{x} \neq \mathbf{0}$, $A\mathbf{x} = 0\mathbf{x}$ Theorem ZSSM [676] $\iff \lambda = 0$ is an eigenvalue of ADefinition EEM [945]

With an equivalence about singular matrices we can update our list of equivalences about nonsingular matrices.

Theorem NME8 Nonsingular Matrix Equivalences, Round 8

Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.
- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.
- 7. The column space of A is \mathbb{C}^n , $\mathcal{C}(A) = \mathbb{C}^n$.
- 8. The columns of A are a basis for \mathbb{C}^n .
- 9. The rank of A is n, r(A) = n.
- 10. The nullity of A is zero, n(A) = 0.

- 11. The determinant of A is nonzero, $det(A) \neq 0$.
- 12. $\lambda = 0$ is not an eigenvalue of A.

Proof The equivalence of the first and last statements is the contrapositive of Theorem SMZE [1003], so we are able to improve on Theorem NME7 [936]. ■

Certain changes to a matrix change its eigenvalues in a predictable way.

Theorem ESMM Eigenvalues of a Scalar Multiple of a Matrix

Suppose A is a square matrix and λ is an eigenvalue of A. Then $\alpha\lambda$ is an eigenvalue of αA .

Proof Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for λ . Then

$\left(\alpha A\right)\mathbf{x}=\alpha\left(A\mathbf{x}\right)$	Theorem MMSMM $[477]$
$= \alpha \left(\lambda \mathbf{x} \right)$	\mathbf{x} eigenvector of A
$= (\alpha \lambda) \mathbf{x}$	Property SMAC [206]

So $\mathbf{x} \neq \mathbf{0}$ is an eigenvector of αA for the eigenvalue $\alpha \lambda$.

Unfortunately, there are not parallel theorems about the sum or product of arbitrary matrices. But we can prove a similar result for powers of a matrix.

 \square

Theorem EOMP Eigenvalues Of Matrix Powers

Suppose A is a square matrix, λ is an eigenvalue of A, and $s \ge 0$ is an integer. Then λ^s is an eigenvalue of A^s .

Proof Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for λ . Suppose A has size n. Then we proceed by induction on s (Technique I [1630]). First, for s = 0,

$A^s \mathbf{x} = A^0 \mathbf{x}$	
$=I_n\mathbf{x}$	
$= \mathbf{x}$	Theorem MMIM [475]
$= 1\mathbf{x}$	Property OC [206]
$=\lambda^0 {f x}$	
$=\lambda^s {f x}$	

so λ^s is an eigenvalue of A^s in this special case. If we assume the theorem is true for s, then we find

$A^{s+1}\mathbf{x} = A^s A \mathbf{x}$	
$=A^{s}\left(\lambda\mathbf{x}\right)$	x eigenvector of A for λ
$=\lambda\left(A^{s}\mathbf{x} ight)$	Theorem MMSMM [477]
$=\lambda\left(\lambda^{s}\mathbf{x} ight)$	Induction hypothesis

 $= (\lambda \lambda^{s}) \mathbf{x}$ Property SMAC [206] $= \lambda^{s+1} \mathbf{x}$

So $\mathbf{x} \neq \mathbf{0}$ is an eigenvector of A^{s+1} for λ^{s+1} , and induction tells us the theorem is true for all $s \ge 0$.

While we cannot prove that the sum of two arbitrary matrices behaves in any reasonable way with regard to eigenvalues, we can work with the sum of dissimilar powers of the *same* matrix. We have already seen two connections between eigenvalues and polynomials, in the proof of Theorem EMHE [953] and the characteristic polynomial (Definition CP [960]). Our next theorem strengthens this connection.

Theorem EPM

Eigenvalues of the Polynomial of a Matrix

Suppose A is a square matrix and λ is an eigenvalue of A. Let q(x) be a polynomial in the variable x. Then $q(\lambda)$ is an eigenvalue of the matrix q(A).

Proof Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for λ , and write $q(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$. Then

$$q(A)\mathbf{x} = (a_0A^0 + a_1A^1 + a_2A^2 + \dots + a_mA^m)\mathbf{x}$$

= $(a_0A^0)\mathbf{x} + (a_1A^1)\mathbf{x} + (a_2A^2)\mathbf{x} + \dots + (a_mA^m)\mathbf{x}$ Theorem MMDAA [476]
= $a_0(A^0\mathbf{x}) + a_1(A^1\mathbf{x}) + a_2(A^2\mathbf{x}) + \dots + a_m(A^m\mathbf{x})$ Theorem MMSMM [477]
= $a_0(\lambda^0\mathbf{x}) + a_1(\lambda^1\mathbf{x}) + a_2(\lambda^2\mathbf{x}) + \dots + a_m(\lambda^m\mathbf{x})$ Theorem EOMP [1006]

$$= (a_0\lambda^0)\mathbf{x} + (a_1\lambda^1)\mathbf{x} + (a_2\lambda^2)\mathbf{x} + \dots + (a_m\lambda^m)\mathbf{x}$$
$$= (a_0\lambda^0 + a_1\lambda^1 + a_2\lambda^2 + \dots + a_m\lambda^m)\mathbf{x}$$
$$= q(\lambda)\mathbf{x}$$

Property SMAC [206] Property DSAC [206]

So $\mathbf{x} \neq 0$ is an eigenvector of q(A) for the eigenvalue $q(\lambda)$.

Example BDE Building desired eigenvalues

In Example ESMS4 [970] the 4×4 symmetric matrix

$$C = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

is shown to have the three eigenvalues $\lambda = 3, 1, -1$. Suppose we wanted a 4×4 matrix that has the three eigenvalues $\lambda = 4, 0, -2$. We can employ Theorem EPM [1007] by finding a polynomial that converts 3 to 4, 1 to 0, and -1 to -2. Such a polynomial is called an **interpolating polynomial**, and in this example we can use

$$r(x) = \frac{1}{4}x^2 + x - \frac{5}{4}$$

We will not discuss how to concoct this polynomial, but a text on numerical analysis should provide the details or see Section CF [1940]. For now, simply verify that r(3) = 4, r(1) = 0 and r(-1) = -2.

Now compute

$$\begin{aligned} r(C) &= \frac{1}{4}C^2 + C - \frac{5}{4}I_4 \\ &= \frac{1}{4} \begin{bmatrix} 3 & 2 & 2 & 2 \\ 2 & 3 & 2 & 2 \\ 2 & 2 & 3 & 2 \\ 2 & 2 & 2 & 3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} - \frac{5}{4} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \\ 3 & 3 & 1 & 1 \\ 3 & 3 & 1 & 1 \end{bmatrix} \end{aligned}$$

Theorem EPM [1007] tells us that if r(x) transforms the eigenvalues in the desired manner, then r(C) will have the desired eigenvalues. You can check this by computing the eigenvalues of r(C) directly. Furthermore, notice that the multiplicities are the same, and the eigenspaces of C and r(C) are identical.

Inverses and transposes also behave predictably with regard to their eigenvalues.

Theorem EIM

Eigenvalues of the Inverse of a Matrix

Suppose A is a square nonsingular matrix and λ is an eigenvalue of A. Then $\frac{1}{\lambda}$ is an eigenvalue of the matrix A^{-1} .

Proof Notice that since A is assumed nonsingular, A^{-1} exists by Theorem NI [542], but

more importantly, $\frac{1}{\lambda}$ does not involve division by zero since Theorem SMZE [1003] prohibits this possibility.

Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for λ . Suppose A has size n. Then

$$A^{-1}\mathbf{x} = A^{-1}(1\mathbf{x}) \qquad \text{Property OC [206]}$$
$$= A^{-1}(\frac{1}{\lambda}\lambda\mathbf{x}) \qquad \text{Property MICN [1598]}$$
$$= \frac{1}{\lambda}A^{-1}(\lambda\mathbf{x}) \qquad \text{Theorem MMSMM [477]}$$
$$= \frac{1}{\lambda}A^{-1}(A\mathbf{x}) \qquad \text{Definition EEM [945]}$$
$$= \frac{1}{\lambda}(A^{-1}A)\mathbf{x} \qquad \text{Theorem MMA [478]}$$
$$= \frac{1}{\lambda}I_n\mathbf{x} \qquad \text{Definition MI [506]}$$
$$= \frac{1}{\lambda}\mathbf{x} \qquad \text{Theorem MMIM [475]}$$

So $\mathbf{x} \neq 0$ is an eigenvector of A^{-1} for the eigenvalue $\frac{1}{\lambda}$.

The theorems above have a similar style to them, a style you should consider using when confronted with a need to prove a theorem about eigenvalues and eigenvectors. So far we have been able to reserve the characteristic polynomial for strictly computational purposes. However, the next theorem, whose statement resembles the preceding theorems, has an easier proof if we employ the characteristic polynomial and results about determinants.

Theorem ETM Eigenvalues of the Transpose of a Matrix

Suppose A is a square matrix and λ is an eigenvalue of A. Then λ is an eigenvalue of the matrix A^t .

Proof Suppose A has size n. Then

$$p_A(x) = \det (A - xI_n)$$
 Definition CP [960]

$$= \det ((A - xI_n)^t)$$
 Theorem DT [899]

$$= \det (A^t - (xI_n)^t)$$
 Theorem TMA [436]

$$= \det (A^t - xI_n^t)$$
 Theorem TMSM [436]

$$= \det (A^t - xI_n)$$
 Definition IM [175]

$$= p_{A^t}(x)$$
 Definition CP [960]

So A and A^t have the same characteristic polynomial, and by Theorem EMRCP [962], their eigenvalues are identical and have equal algebraic multiplicities. Notice that what we have proved here is a bit stronger than the stated conclusion in the theorem.

If a matrix has only real entries, then the computation of the characteristic polynomial (Definition CP [960]) will result in a polynomial with coefficients that are real numbers. Complex numbers could result as roots of this polynomial, but they are roots of quadratic factors with real coefficients, and as such, come in conjugate pairs. The next theorem proves this, and a bit more, without mentioning the characteristic polynomial.

Theorem ERMCP Eigenvalues of Real Matrices come in Conjugate Pairs

Suppose A is a square matrix with real entries and \mathbf{x} is an eigenvector of A for the eigenvalue λ . Then $\overline{\mathbf{x}}$ is an eigenvector of A for the eigenvalue $\overline{\lambda}$.

Proof

$$A\overline{\mathbf{x}} = \overline{A}\overline{\mathbf{x}}$$
A has real entries $= \overline{A}\overline{\mathbf{x}}$ Theorem MMCC [481] $= \overline{\lambda}\overline{\mathbf{x}}$ \mathbf{x} eigenvector of A $= \overline{\lambda}\overline{\mathbf{x}}$ Theorem CRSM [397]

So $\overline{\mathbf{x}}$ is an eigenvector of A for the eigenvalue $\overline{\lambda}$.

This phenomenon is amply illustrated in Example CEMS6 [974], where the four complex eigenvalues come in two pairs, and the two basis vectors of the eigenspaces are complex conjugates of each other. Theorem ERMCP [1012] can be a time-saver for computing eigenvalues and eigenvectors of real matrices with complex eigenvalues, since the conjugate eigenvalue and eigenspace can be inferred from the theorem rather than computed.

Subsection ME Multiplicities of Eigenvalues

A polynomial of degree n will have exactly n roots. From this fact about polynomial equations we can say more about the algebraic multiplicities of eigenvalues.

Theorem DCP

Degree of the Characteristic Polynomial

Suppose that A is a square matrix of size n. Then the characteristic polynomial of A, $p_A(x)$, has degree n.

Proof We will prove a more general result by induction (Technique I [1630]). Then the theorem will be true as a special case. We will carefully state this result as a proposition indexed by $m, m \ge 1$.

P(m): Suppose that A is an $m \times m$ matrix whose entries are complex numbers or linear polynomials in the variable x of the form c - x, where c is a complex number. Suppose further that there are exactly k entries that contain x and that no row or column contains more than one such entry. Then, when k = m, det (A) is a polynomial in x of degree m, with leading coefficient ± 1 , and when k < m, det (A) is a polynomial in x of degree k or less.

Base Case: Suppose A is a 1×1 matrix. Then its determinant is equal to the lone entry (Definition DM [894]). When k = m = 1, the entry is of the form c - x, a polynomial in x of

degree m = 1 with leading coefficient -1. When k < m, then k = 0 and the entry is simply a complex number, a polynomial of degree $0 \le k$. So P(1) is true.

Induction Step: Assume P(m) is true, and that A is an $(m+1) \times (m+1)$ matrix with k entries of the form c - x. There are two cases to consider.

Suppose k = m + 1. Then every row and every column will contain an entry of the form c - x. Suppose that for the first row, this entry is in column t. Compute the determinant of A by an expansion about this first row (Definition DM [894]). The term associated with entry t of this row will be of the form

$$(c-x)(-1)^{1+t} \det (A(1|t))$$

The submatrix A(1|t) is an $m \times m$ matrix with k = m terms of the form c - x, no more than one per row or column. By the induction hypothesis, det (A(1|t)) will be a polynomial in xof degree m with coefficient ± 1 . So this entire term is then a polynomial of degree m + 1with leading coefficient ± 1 .

The remaining terms (which constitute the sum that is the determinant of A) are products of complex numbers from the first row with cofactors built from submatrices that lack the first row of A and lack some column of A, other than column t. As such, these submatrices are $m \times m$ matrices with k = m - 1 < m entries of the form c - x, no more than one per row or column. Applying the induction hypothesis, we see that these terms are polynomials in x of degree m - 1 or less. Adding the single term from the entry in column t with all these others, we see that det (A) is a polynomial in x of degree m + 1 and leading coefficient ± 1 .

The second case occurs when k < m + 1. Now there is a row of A that does not contain an entry of the form c - x. We consider the determinant of A by expanding about this row (Theorem DER [897]), whose entries are all complex numbers. The cofactors employed are built from submatrices that are $m \times m$ matrices with either k or k-1 entries of the form c-x, no more than one per row or column. In either case, $k \leq m$, and we can apply the induction hypothesis to see that the determinants computed for the cofactors are all polynomials of degree k or less. Summing these contributions to the determinant of A yields a polynomial in x of degree k or less, as desired.

Definition CP [960] tells us that the characteristic polynomial of an $n \times n$ matrix is the determinant of a matrix having exactly n entries of the form c - x, no more than one per row or column. As such we can apply P(n) to see that the characteristic polynomial has degree n.

Theorem NEM Number of Eigenvalues of a Matrix

Suppose that A is a square matrix of size n with distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$. Then

$$\sum_{i=1}^{k} \alpha_A(\lambda_i) = n$$

Proof By the definition of the algebraic multiplicity (Definition AME [968]), we can factor the characteristic polynomial as

$$p_A(x) = c(x - \lambda_1)^{\alpha_A(\lambda_1)} (x - \lambda_2)^{\alpha_A(\lambda_2)} (x - \lambda_3)^{\alpha_A(\lambda_3)} \cdots (x - \lambda_k)^{\alpha_A(\lambda_k)}$$

 \square

where c is a nonzero constant. (We could prove that $c = (-1)^n$, but we do not need that specificity right now. See Exercise PEE.T30 [1024]) The left-hand side is a polynomial of degree n by Theorem DCP [1013] and the right-hand side is a polynomial of degree $\sum_{i=1}^k \alpha_A(\lambda_i)$. So the equality of the polynomials' degrees gives the equality $\sum_{i=1}^k \alpha_A(\lambda_i) = n$.

Theorem ME Multiplicities of an Eigenvalue

Suppose that A is a square matrix of size n and λ is an eigenvalue. Then

$$1 \le \gamma_A(\lambda) \le \alpha_A(\lambda) \le n$$

Proof Since λ is an eigenvalue of A, there is an eigenvector of A for λ , **x**. Then $\mathbf{x} \in \mathcal{E}_A(\lambda)$, so $\gamma_A(\lambda) \ge 1$, since we can extend $\{\mathbf{x}\}$ into a basis of $\mathcal{E}_A(\lambda)$ (Theorem ELIS [850]).

To show that $\gamma_A(\lambda) \leq \alpha_A(\lambda)$ is the most involved portion of this proof. To this end, let $g = \gamma_A(\lambda)$ and let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_g$ be a basis for the eigenspace of λ , $\mathcal{E}_A(\lambda)$. Construct another n - g vectors, $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{n-g}$, so that

$$\{\mathbf{x}_1,\,\mathbf{x}_2,\,\mathbf{x}_3,\,\ldots,\,\mathbf{x}_g,\,\mathbf{y}_1,\,\mathbf{y}_2,\,\mathbf{y}_3,\,\ldots,\,\mathbf{y}_{n-g}\}$$

is a basis of \mathbb{C}^n . This can be done by repeated applications of Theorem ELIS [850]. Finally, define a matrix S by

$$S = [\mathbf{x}_1 | \mathbf{x}_2 | \mathbf{x}_3 | \dots | \mathbf{x}_g | \mathbf{y}_1 | \mathbf{y}_2 | \mathbf{y}_3 | \dots | \mathbf{y}_{n-g}] = [\mathbf{x}_1 | \mathbf{x}_2 | \mathbf{x}_3 | \dots | \mathbf{x}_g | R]$$

 \square

where R is an $n \times (n - g)$ matrix whose columns are $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{n-g}$. The columns of S are linearly independent by design, so S is nonsingular (Theorem NMLIC [330]) and therefore invertible (Theorem NI [542]). Then,

$$\begin{aligned} \mathbf{[e_1|e_2|e_3|\dots|e_n]} &= I_n \\ &= S^{-1}S \\ &= S^{-1}[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\dots|\mathbf{x}_g|R] \\ &= [S^{-1}\mathbf{x}_1|S^{-1}\mathbf{x}_2|S^{-1}\mathbf{x}_3|\dots|S^{-1}\mathbf{x}_g|S^{-1}R \end{aligned}$$

 So

$$S^{-1}\mathbf{x}_i = \mathbf{e}_i \quad 1 \le i \le g \tag{(*)}$$

Preparations in place, we compute the characteristic polynomial of A,

$$p_A(x) = \det (A - xI_n)$$

= 1 det $(A - xI_n)$
= det (I_n) det $(A - xI_n)$
= det $(S^{-1}S)$ det $(A - xI_n)$
= det (S^{-1}) det (S) det $(A - xI_n)$
= det (S^{-1}) det $(A - xI_n)$ det (S)
= det $(S^{-1}(A - xI_n)S)$
= det $(S^{-1}AS - S^{-1}xI_nS)$

Definition CP [960] Property OCN [1597] Definition DM [894] Definition MI [506] Theorem DRMM [938] Property CMCN [1597] Theorem DRMM [938] Theorem MMDAA [476]

Definition MI [506] Definition CP [960]

Theorem MMSMM [477] Theorem MMIM [475]

$$= \det \left(S^{-1}AS - xS^{-1}I_nS \right)$$
$$= \det \left(S^{-1}AS - xS^{-1}S \right)$$
$$= \det \left(S^{-1}AS - xI_n \right)$$
$$= p_{S^{-1}AS} \left(x \right)$$

What can we learn then about the matrix $S^{-1}AS$?

$$\begin{split} S^{-1}AS &= S^{-1}A[\mathbf{x}_{1}|\mathbf{x}_{2}|\mathbf{x}_{3}|\dots|\mathbf{x}_{g}|R] \\ &= S^{-1}[A\mathbf{x}_{1}|A\mathbf{x}_{2}|A\mathbf{x}_{3}|\dots|A\mathbf{x}_{g}|AR] \\ &= S^{-1}[\lambda\mathbf{x}_{1}|\lambda\mathbf{x}_{2}|\lambda\mathbf{x}_{3}|\dots|\lambda\mathbf{x}_{g}|AR] \\ &= [S^{-1}\lambda\mathbf{x}_{1}|S^{-1}\lambda\mathbf{x}_{2}|S^{-1}\lambda\mathbf{x}_{3}|\dots|S^{-1}\lambda\mathbf{x}_{g}|S^{-1}AR] \\ &= [\lambda S^{-1}\mathbf{x}_{1}|\lambda S^{-1}\mathbf{x}_{2}|\lambda S^{-1}\mathbf{x}_{3}|\dots|AS^{-1}\mathbf{x}_{g}|S^{-1}AR] \\ &= [\lambda e_{1}|\lambda e_{2}|\lambda e_{3}|\dots|\lambda e_{g}|S^{-1}AR] \\ \end{split}$$
 Definition MM [468]
 Theorem MMSMM [477]
 S^{-1}S = I_{n}, ((*) above) \end{split}

Now imagine computing the characteristic polynomial of A by computing the characteristic polynomial of $S^{-1}AS$ using the form just obtained. The first g columns of $S^{-1}AS$ are all zero, save for a λ on the diagonal. So if we compute the determinant by expanding about the first column, successively, we will get successive factors of $(\lambda - x)$. More precisely, let T be the square matrix of size n - g that is formed from the last n - g rows and last n - g columns of $S^{-1}AR$. Then

$$p_A(x) = p_{S^{-1}AS}(x) = (\lambda - x)^g p_T(x).$$

This says that $(x - \lambda)$ is a factor of the characteristic polynomial *at least* g times, so the algebraic multiplicity of λ as an eigenvalue of A is greater than or equal to g (Definition AME [968]). In other words,

$$\gamma_A(\lambda) = g \le \alpha_A(\lambda)$$

as desired.

Theorem NEM [1015] says that the sum of the algebraic multiplicities for *all* the eigenvalues of A is equal to n. Since the algebraic multiplicity is a positive quantity, no single algebraic multiplicity can exceed n without the sum of all of the algebraic multiplicities doing the same.

Theorem MNEM

Maximum Number of Eigenvalues of a Matrix

Suppose that A is a square matrix of size n. Then A cannot have more than n distinct eigenvalues. \Box

Proof Suppose that A has k distinct eigenvalues, $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$. Then

$$k = \sum_{i=1}^{k} 1$$

$$\leq \sum_{i=1}^{k} \alpha_A(\lambda_i)$$
 Theorem ME [1016]

$$= n$$
 Theorem NEM [1015]

Subsection EHM Eigenvalues of Hermitian Matrices

Recall that a matrix is Hermitian (or self-adjoint) if $A = A^*$ (Definition HM [486]). In the case where A is a matrix whose entries are all real numbers, being Hermitian is identical to being symmetric (Definition SYM [434]). Keep this in mind as you read the next two theorems. Their hypotheses could be changed to "suppose A is a real symmetric matrix."

Theorem HMRE Hermitian Matrices have Real Eigenvalues

Suppose that A is a Hermitian matrix and λ is an eigenvalue of A. Then $\lambda \in \mathbb{R}$.

Proof Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for the eigenvalue λ . Then by Theorem PIP [406] we know $\langle \mathbf{x}, \mathbf{x} \rangle \neq 0$. So

 $\lambda = \frac{1}{\langle \mathbf{x}, \, \mathbf{x} \rangle} \lambda \, \langle \mathbf{x}, \, \mathbf{x} \rangle$ $= \frac{1}{\langle \mathbf{x}, \, \mathbf{x} \rangle} \, \langle \mathbf{x}, \, \lambda \mathbf{x} \rangle$ $= \frac{1}{\langle \mathbf{x}, \, \mathbf{x} \rangle} \, \langle \mathbf{x}, \, A \mathbf{x} \rangle$

Property MICN [1598]

Theorem IPSM [401]

Definition EEM [945]

 \square

$=rac{1}{\left< {f x},{f x} ight>} \left< A{f x},{f x} ight>$	Theorem HMIP [486]
$=rac{1}{\left\langle \mathbf{x},\mathbf{x} ight angle }\left\langle \lambda\mathbf{x},\mathbf{x} ight angle$	Definition EEM [945]
$=rac{1}{\langle {f x},{f x} angle }\overline{\lambda}\left\langle {f x},{f x} ight angle$	Theorem IPSM [401]
$=\overline{\lambda}$	Property MICN [1598]

If a complex number is equal to its conjugate, then it has a complex part equal to zero, and therefore is a real number.

Notice the appealing symmetry to the justifications given for the steps of this proof. In the center is the ability to pitch a Hermitian matrix from one side of the inner product to the other.

Look back and compare Example ESMS4 [970] and Example CEMS6 [974]. In Example CEMS6 [974] the matrix has only real entries, yet the characteristic polynomial has roots that are complex numbers, and so the matrix has complex eigenvalues. However, in Example ESMS4 [970], the matrix has only real entries, but is also symmetric, and hence Hermitian. So by Theorem HMRE [1020], we were guaranteed eigenvalues that are real numbers.

In many physical problems, a matrix of interest will be real and symmetric, or Hermitian. Then if the eigenvalues are to represent physical quantities of interest, Theorem HMRE [1020] guarantees that these values will not be complex numbers.

The eigenvectors of a Hermitian matrix also enjoy a pleasing property that we will exploit later.

Theorem HMOE Hermitian Matrices have Orthogonal Eigenvectors

Suppose that A is a Hermitian matrix and \mathbf{x} and \mathbf{y} are two eigenvectors of A for different eigenvalues. Then \mathbf{x} and \mathbf{y} are orthogonal vectors.

Proof Let \mathbf{x} be an eigenvector of A for λ and let \mathbf{y} be an eigenvector of A for a different eigenvalue ρ . So we have $\lambda - \rho \neq 0$. Then

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{\lambda - \rho} (\lambda - \rho) \langle \mathbf{x}, \mathbf{y} \rangle$$
Property MICN [1598]
$$= \frac{1}{\lambda - \rho} (\lambda \langle \mathbf{x}, \mathbf{y} \rangle - \rho \langle \mathbf{x}, \mathbf{y} \rangle)$$
Property MICN [1598]
$$= \frac{1}{\lambda - \rho} (\langle \overline{\lambda} \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{x}, \rho \mathbf{y} \rangle)$$
Theorem IPSM [401]
$$= \frac{1}{\lambda - \rho} (\langle \lambda \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{x}, \rho \mathbf{y} \rangle)$$
Theorem HMRE [1020
$$= \frac{1}{\lambda - \rho} (\langle A\mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{x}, A\mathbf{y} \rangle)$$
Definition EEM [945]
$$= \frac{1}{\lambda - \rho} (\langle A\mathbf{x}, \mathbf{y} \rangle - \langle A\mathbf{x}, \mathbf{y} \rangle)$$
Theorem HMIP [486]
$$= \frac{1}{\lambda - \rho} (0)$$
Property AICN [1597]
$$= 0$$

This equality says that \mathbf{x} and \mathbf{y} are orthogonal vectors (Definition OV [408]).

Notice again how the key step in this proof is the fundamental property of a Hermitian matrix (Theorem HMIP [486]) — the ability to swap A across the two arguments of the inner product. We'll build on these results and continue to see some more interesting properties in Section OD [1414].

Subsection READ Reading Questions

- 1. How can you identify a nonsingular matrix just by looking at its eigenvalues?
- 2. How many different eigenvalues may a square matrix of size n have?
- 3. What is amazing about the eigenvalues of a Hermitian matrix and why is it amazing?

Subsection EXC Exercises

T10 Suppose that A is a square matrix. Prove that the constant term of the characteristic polynomial of A is equal to the determinant of A. Contributed by Robert Beezer Solution [1026]

T20 Suppose that A is a square matrix. Prove that a single vector may not be an eigenvector of A for two different eigenvalues. Contributed by Robert Beezer Solution [1026]

T22 Suppose that U is a unitary matrix with eigenvalue λ . Prove that λ had modulus 1, i.e. $|\lambda| = 1$. This says that all of the eigenvalues of a unitary matrix lie on the unit circle of the complex plane.

Contributed by Robert Beezer

T30 Theorem DCP [1013] tells us that the characteristic polynomial of a square matrix of size n has degree n. By suitably augmenting the proof of Theorem DCP [1013] prove that the coefficient of x^n in the characteristic polynomial is $(-1)^n$. Contributed by Robert Beezer

T50 Theorem EIM [1009] says that if λ is an eigenvalue of the nonsingular matrix A, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} . Write an alternate proof of this theorem using the characteristic

polynomial and without making reference to an eigenvector of A for λ . Contributed by Robert Beezer Solution [1027]

Subsection SOL Solutions

T10 Contributed by Robert Beezer Statement [1024] Suppose that the characteristic polynomial of A is

$$p_A(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

Then

$$a_{0} = a_{0} + a_{1}(0) + a_{2}(0)^{2} + \dots + a_{n}(0)^{n}$$

= $p_{A}(0)$
= $\det (A - 0I_{n})$ Definition CP [960]
= $\det (A)$

T20 Contributed by Robert Beezer Statement [1024] Suppose that the vector $\mathbf{x} \neq \mathbf{0}$ is an eigenvector of A for the two eigenvalues λ and ρ , where $\lambda \neq \rho$. Then $\lambda - \rho \neq 0$, and we also have

$0 = A\mathbf{x} - A\mathbf{x}$	Property AIC $[206]$
$= \lambda \mathbf{x} - \rho \mathbf{x}$	Definition EEM [945]

 $= (\lambda - \rho)\mathbf{x}$ Property DSAC [206]

By Theorem SMEZV [679], either $\lambda - \rho = 0$ or $\mathbf{x} = \mathbf{0}$, which are both contradictions.

T50 Contributed by Robert Beezer Statement [1024] Since λ is an eigenvalue of a nonsingular matrix, $\lambda \neq 0$ (Theorem SMZE [1003]). A is invertible (Theorem NI [542]), and so $-\lambda A$ is invertible (Theorem MISM [523]). Thus $-\lambda A$ is nonsingular (Theorem NI [542]) and det $(-\lambda A) \neq 0$ (Theorem SMZD [935]).

$$p_{A^{-1}}\left(\frac{1}{\lambda}\right) = \det\left(A^{-1} - \frac{1}{\lambda}I_n\right) \qquad \text{Definition CP [960]}$$

$$= 1 \det\left(A^{-1} - \frac{1}{\lambda}I_n\right) \qquad \text{Property OCN [1597]}$$

$$= \frac{1}{\det\left(-\lambda A\right)} \det\left(-\lambda A\right) \det\left(A^{-1} - \frac{1}{\lambda}I_n\right) \qquad \text{Property MICN [1598]}$$

$$= \frac{1}{\det\left(-\lambda A\right)} \det\left(\left(-\lambda A\right)\left(A^{-1} - \frac{1}{\lambda}I_n\right)\right) \qquad \text{Theorem DRMM [938]}$$

$$= \frac{1}{\det\left(-\lambda A\right)} \det\left(-\lambda AA^{-1} - (-\lambda A)\frac{1}{\lambda}I_n\right) \qquad \text{Theorem MMDAA [476]}$$

$$= \frac{1}{\det\left(-\lambda A\right)} \det\left(-\lambda I_n - (-\lambda A)\frac{1}{\lambda}I_n\right) \qquad \text{Definition MI [506]}$$

$$= \frac{1}{\det\left(-\lambda A\right)} \det\left(-\lambda I_n + \lambda\frac{1}{\lambda}AI_n\right) \qquad \text{Theorem MMSMM [477]}$$

$$= \frac{1}{\det(-\lambda A)} \det(-\lambda I_n + 1AI_n)$$
Property MICN [1598]

$$= \frac{1}{\det(-\lambda A)} \det(-\lambda I_n + AI_n)$$
Property OCN [1597]

$$= \frac{1}{\det(-\lambda A)} \det(-\lambda I_n + A)$$
Theorem MMIM [475]

$$= \frac{1}{\det(-\lambda A)} \det(A - \lambda I_n)$$
Property ACM [430]

$$= \frac{1}{\det(-\lambda A)} p_A(\lambda)$$
Definition CP [960]

$$= \frac{1}{\det(-\lambda A)} 0$$
Theorem EMRCP [962]

$$= 0$$
Property ZCN [1597]

So $\frac{1}{\lambda}$ is a root of the characteristic polynomial of A^{-1} and so is an eigenvalue of A^{-1} . This proof is due to Sara Bucht.

Section SD Similarity and Diagonalization

This section's topic will perhaps seem out of place at first, but we will make the connection soon with eigenvalues and eigenvectors. This is also our first look at one of the central ideas of Chapter R [1260].

Subsection SM Similar Matrices

The notion of matrices being "similar" is a lot like saying two matrices are row-equivalent. Two similar matrices are not equal, but they share many important properties. This section, and later sections in Chapter R [1260] will be devoted in part to discovering just what these common properties are.

First, the main definition for this section.

Definition SIM Similar Matrices

Suppose A and B are two square matrices of size n. Then A and B are similar if there

exists a nonsingular matrix of size n, S, such that $A = S^{-1}BS$.

We will say "A is similar to B via S" when we want to emphasize the role of S in the relationship between A and B. Also, it doesn't matter if we say A is similar to B, or B is similar to A. If one statement is true then so is the other, as can be seen by using S^{-1} in place of S (see Theorem SER [1033] for the careful proof). Finally, we will refer to $S^{-1}BS$ as a **similarity transformation** when we want to emphasize the way S changes B. OK, enough about language, let's build a few examples.

Example SMS5 Similar matrices of size 5

If you wondered if there are examples of similar matrices, then it won't be hard to convince you they exist. Define

$$B = \begin{bmatrix} -4 & 1 & -3 & -2 & 2\\ 1 & 2 & -1 & 3 & -2\\ -4 & 1 & 3 & 2 & 2\\ -3 & 4 & -2 & -1 & -3\\ 3 & 1 & -1 & 1 & -4 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 2 & -1 & 1 & 1\\ 0 & 1 & -1 & -2 & -1\\ 1 & 3 & -1 & 1 & 1\\ -2 & -3 & 3 & 1 & -2\\ 1 & 3 & -1 & 2 & 1 \end{bmatrix}$$

Check that S is nonsingular and then compute

 $A = S^{-1}BS$

 \triangle

$$= \begin{bmatrix} 10 & 1 & 0 & 2 & -5 \\ -1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 2 & 1 & -3 \\ 0 & 0 & -1 & 0 & 1 \\ -4 & -1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} -4 & 1 & -3 & -2 & 2 \\ 1 & 2 & -1 & 3 & -2 \\ -4 & 1 & 3 & 2 & 2 \\ -3 & 4 & -2 & -1 & -3 \\ 3 & 1 & -1 & 1 & -4 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 1 & 1 \\ 0 & 1 & -1 & -2 & -1 \\ 1 & 3 & -1 & 1 & 1 \\ -2 & -3 & 3 & 1 & -2 \\ 1 & 3 & -1 & 2 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -10 & -27 & -29 & -80 & -25 \\ -2 & 6 & 6 & 10 & -2 \\ -3 & 11 & -9 & -14 & -9 \\ -1 & -13 & 0 & -10 & -1 \\ 11 & 35 & 6 & 49 & 19 \end{bmatrix}$$

So by this construction, we know that A and B are similar.

Let's do that again.

Example SMS3 Similar matrices of size 3 Define

$$B = \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \qquad \qquad S = \begin{bmatrix} 1 & 1 & 2 \\ -2 & -1 & -3 \\ 1 & -2 & 0 \end{bmatrix}$$

Check that S is nonsingular and then compute

$$A = S^{-1}BS$$

 \boxtimes

$$= \begin{bmatrix} -6 & -4 & -1 \\ -3 & -2 & -1 \\ 5 & 3 & 1 \end{bmatrix} \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ -2 & -1 & -3 \\ 1 & -2 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

So by this construction, we know that A and B are similar. But before we move on, look at how pleasing the form of A is. Not convinced? Then consider that several computations related to A are especially easy. For example, in the spirit of Example DUTM [903], det (A) =(-1)(3)(-1) = 3. Similarly, the characteristic polynomial is straightforward to compute by hand, $p_A(x) = (-1 - x)(3 - x)(-1 - x) = -(x - 3)(x + 1)^2$ and since the result is already factored, the eigenvalues are transparently $\lambda = 3$, -1. Finally, the eigenvectors of A are just the standard unit vectors (Definition SUV [409]).

Subsection PSM Properties of Similar Matrices

Similar matrices share many properties and it is these theorems that justify the choice of the word "similar." First we will show that similarity is an **equivalence relation**. Equivalence relations are important in the study of various algebras and can always be regarded as a kind

of weak version of equality. Sort of alike, but not quite equal. The notion of two matrices being row-equivalent is an example of an equivalence relation we have been working with since the beginning of the course (see Exercise RREF.T11 [95]). Row-equivalent matrices are not equal, but they are a lot alike. For example, row-equivalent matrices have the same rank. Formally, an equivalence relation requires three conditions hold: reflexive, symmetric and transitive. We will illustrate these as we prove that similarity is an equivalence relation.

Theorem SER Similarity is an Equivalence Relation

Suppose A, B and C are square matrices of size n. Then

- 1. A is similar to A. (Reflexive)
- 2. If A is similar to B, then B is similar to A. (Symmetric)
- 3. If A is similar to B and B is similar to C, then A is similar to C. (Transitive)

Proof To see that A is similar to A, we need only demonstrate a nonsingular matrix that effects a similarity transformation of A to A. I_n is nonsingular (since it row-reduces to the identity matrix, Theorem NMRRI [176]), and

$$I_n^{-1}AI_n = I_nAI_n = A$$

If we assume that A is similar to B, then we know there is a nonsingular matrix S so that $A = S^{-1}BS$ by Definition SIM [1029]. By Theorem MIMI [521], S^{-1} is invertible, and by

Theorem NI [542] is therefore nonsingular. So

$$(S^{-1})^{-1}A(S^{-1}) = SAS^{-1}$$
$$= SS^{-1}BSS^{-1}$$
$$= (SS^{-1}) B (SS^{-1})$$
$$= I_n BI_n$$
$$= B$$

Theorem MIMI [521] Definition SIM [1029] Theorem MMA [478] Definition MI [506] Theorem MMIM [475]

and we see that B is similar to A.

Assume that A is similar to B, and B is similar to C. This gives us the existence of two nonsingular matrices, S and R, such that $A = S^{-1}BS$ and $B = R^{-1}CR$, by Definition SIM [1029]. (Notice how we have to assume $S \neq R$, as will usually be the case.) Since S and R are invertible, so too RS is invertible by Theorem SS [520] and then nonsingular by Theorem NI [542]. Now

$$(RS)^{-1}C(RS) = S^{-1}R^{-1}CRS$$

$$= S^{-1} (R^{-1}CR) S$$

$$= S^{-1}BS$$

$$= A$$
Theorem SS [520]
Theorem MMA [478]
Definition SIM [1029]

so A is similar to C via the nonsingular matrix RS.

Here's another theorem that tells us exactly what sorts of properties similar matrices share.

Theorem SMEE

Similar Matrices have Equal Eigenvalues

Suppose A and B are similar matrices. Then the characteristic polynomials of A and B are equal, that is, $p_A(x) = p_B(x)$.

Proof Let *n* denote the size of *A* and *B*. Since *A* and *B* are similar, there exists a nonsingular matrix *S*, such that $A = S^{-1}BS$ (Definition SIM [1029]). Then

$$p_A(x) = \det (A - xI_n)$$

$$= \det (S^{-1}BS - xI_n)$$

$$= \det (S^{-1}BS - xS^{-1}I_nS)$$

$$= \det (S^{-1}BS - S^{-1}xI_nS)$$

$$= \det (S^{-1} (B - xI_n) S)$$

$$= \det (S^{-1}) \det (B - xI_n) \det (S)$$

$$= \det (S^{-1}) \det (S) \det (B - xI_n)$$

$$= \det (S^{-1}S) \det (B - xI_n)$$

$$= \det (I_n) \det (B - xI_n)$$

$$= 1 \det (B - xI_n)$$

$$= p_B(x)$$

Definition CP [960] Definition SIM [1029] Theorem MMIM [475] Theorem MMSMM [477] Theorem MMDAA [476] Theorem DRMM [938] Property CMCN [1597] Theorem DRMM [938] Definition MI [506] Definition DM [894] Definition CP [960]

So similar matrices not only have the same set of eigenvalues, the algebraic multiplici-

ties of these eigenvalues will also be the same. However, be careful with this theorem. It is tempting to think the converse is true, and argue that if two matrices have the same eigenvalues, then they are similar. Not so, as the following example illustrates.

Example EENS Equal eigenvalues, not similar Define

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and check that

$$p_A(x) = p_B(x) = 1 - 2x + x^2 = (x - 1)^2$$

and so A and B have equal characteristic polynomials. If the converse of Theorem SMEE [1035] were true, then A and B would be similar. Suppose this is the case. More precisely, suppose there is a nonsingular matrix S so that $A = S^{-1}BS$. Then

$$A = S^{-1}BS = S^{-1}I_2S = S^{-1}S = I_2$$

Clearly $A \neq I_2$ and this contradiction tells us that the converse of Theorem SMEE [1035] is false.

Subsection D Diagonalization

Good things happen when a matrix is similar to a diagonal matrix. For example, the eigenvalues of the matrix are the entries on the diagonal of the diagonal matrix. And it can be a much simpler matter to compute high powers of the matrix. Diagonalizable matrices are also of interest in more abstract settings. Here are the relevant definitions, then our main theorem for this section.

Definition DIM Diagonal Matrix

Suppose that A is a square matrix. Then A is a **diagonal matrix** if $[A]_{ij} = 0$ whenever $i \neq j$.

Definition DZM Diagonalizable Matrix

Suppose A is a square matrix. Then A is **diagonalizable** if A is similar to a diagonal matrix.

 \wedge

Example DAB Diagonalization of Archetype B Archetype B [1652] has a 3×3 coefficient matrix

$$B = \begin{bmatrix} -7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4 \end{bmatrix}$$

and is similar to a diagonal matrix, as can be seen by the following computation with the nonsingular matrix S,

$$S^{-1}BS = \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & -1 & -1 \\ 2 & 3 & 1 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

 \boxtimes

Example SMS3 [1031] provides yet another example of a matrix that is subjected to a similarity transformation and the result is a diagonal matrix. Alright, just how would we find the magic matrix S that can be used in a similarity transformation to produce a diagonal

matrix? Before you read the statement of the next theorem, you might study the eigenvalues and eigenvectors of Archetype B [1652] and compute the eigenvalues and eigenvectors of the matrix in Example SMS3 [1031].

Theorem DC Diagonalization Characterization

Suppose A is a square matrix of size n. Then A is diagonalizable if and only if there exists a linearly independent set S that contains n eigenvectors of A. \Box

Proof (\Leftarrow) Let $S = {\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n}$ be a linearly independent set of eigenvectors of A for the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$. Recall Definition SUV [409] and define

$$R = [\mathbf{x}_{1} | \mathbf{x}_{2} | \mathbf{x}_{3} | \dots | \mathbf{x}_{n}]$$

$$D = \begin{bmatrix} \lambda_{1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{2} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_{3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{n} \end{bmatrix} = [\lambda_{1} \mathbf{e}_{1} | \lambda_{2} \mathbf{e}_{2} | \lambda_{3} \mathbf{e}_{3} | \dots | \lambda_{n} \mathbf{e}_{n}]$$

The columns of R are the vectors of the linearly independent set S and so by Theorem NMLIC [330] the matrix R is nonsingular. By Theorem NI [542] we know R^{-1} exists.

$$R^{-1}AR = R^{-1}A[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\dots|\mathbf{x}_n]$$

= $R^{-1}[A\mathbf{x}_1|A\mathbf{x}_2|A\mathbf{x}_3|\dots|A\mathbf{x}_n]$ Definition MM [468]

$$= R^{-1}[\lambda_{1}\mathbf{x}_{1}|\lambda_{2}\mathbf{x}_{2}|\lambda_{3}\mathbf{x}_{3}|\dots|\lambda_{n}\mathbf{x}_{n}]$$
 Definition EEM [945]

$$= R^{-1}[\lambda_{1}R\mathbf{e}_{1}|\lambda_{2}R\mathbf{e}_{2}|\lambda_{3}R\mathbf{e}_{3}|\dots|\lambda_{n}R\mathbf{e}_{n}]$$
 Definition MVP [461]

$$= R^{-1}[R(\lambda_{1}\mathbf{e}_{1})|R(\lambda_{2}\mathbf{e}_{2})|R(\lambda_{3}\mathbf{e}_{3})|\dots|R(\lambda_{n}\mathbf{e}_{n})]$$
 Theorem MMSMM [477]

$$= R^{-1}R[\lambda_{1}\mathbf{e}_{1}|\lambda_{2}\mathbf{e}_{2}|\lambda_{3}\mathbf{e}_{3}|\dots|\lambda_{n}\mathbf{e}_{n}]$$
 Definition MM [468]

$$= I_{n}D$$
 Definition MI [506]

$$= D$$
 Theorem MMIM [475]

This says that A is similar to the diagonal matrix D via the nonsingular matrix R. Thus A is diagonalizable (Definition DZM [1037]).

 (\Rightarrow) Suppose that A is diagonalizable, so there is a nonsingular matrix of size n

 $T = [\mathbf{y}_1 | \mathbf{y}_2 | \mathbf{y}_3 | \dots | \mathbf{y}_n]$

and a diagonal matrix (recall Definition SUV [409])

$$E = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix} = [d_1 \mathbf{e}_1 | d_2 \mathbf{e}_2 | d_3 \mathbf{e}_3 | \dots | d_n \mathbf{e}_n]$$

such that $T^{-1}AT = E$. Then consider,

$$\begin{split} [A\mathbf{y}_1|A\mathbf{y}_2|A\mathbf{y}_3|\dots|A\mathbf{y}_n] &= A\left[\mathbf{y}_1|\mathbf{y}_2|\mathbf{y}_3|\dots|\mathbf{y}_n\right] & \text{Definition MM [468]} \\ &= AT \\ &= I_nAT & \text{Theorem MMIM [475]} \\ &= TT^{-1}AT & \text{Definition MI [506]} \\ &= TE \\ &= T[d_1\mathbf{e}_1|d_2\mathbf{e}_2|d_3\mathbf{e}_3|\dots|d_n\mathbf{e}_n] \\ &= [T(d_1\mathbf{e}_1)|T(d_2\mathbf{e}_2)|T(d_3\mathbf{e}_3)|\dots|T(d_n\mathbf{e}_n)] & \text{Definition MM [468]} \\ &= [d_1T\mathbf{e}_1|d_2T\mathbf{e}_2|d_3T\mathbf{e}_3|\dots|d_nT\mathbf{e}_n] & \text{Definition MM [468]} \\ &= [d_1\mathbf{y}_1|d_2\mathbf{y}_2|d_3\mathbf{y}_3|\dots|d_n\mathbf{y}_n] & \text{Definition MV P [461]} \end{split}$$

This equality of matrices (Definition ME [427]) allows us to conclude that the individual columns are equal vectors (Definition CVE [199]). That is, $A\mathbf{y}_i = d_i\mathbf{y}_i$ for $1 \le i \le n$. In other words, \mathbf{y}_i is an eigenvector of A for the eigenvalue d_i , $1 \le i \le n$. (Why can't $\mathbf{y}_i = \mathbf{0}$?). Because T is nonsingular, the set containing T's columns, $S = {\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_n}$, is a linearly independent set (Theorem NMLIC [330]). So the set S has all the required properties.

Notice that the proof of Theorem DC [1039] is constructive. To diagonalize a matrix, we need only locate n linearly independent eigenvectors. Then we can construct a nonsingular matrix using the eigenvectors as columns (R) so that $R^{-1}AR$ is a diagonal matrix (D). The entries on the diagonal of D will be the eigenvalues of the eigenvectors used to create R, in

the same order as the eigenvectors appear in R. We illustrate this by **diagonalizing** some matrices.

Example DMS3 Diagonalizing a matrix of size 3

Consider the matrix

$$F = \begin{bmatrix} -13 & -8 & -4\\ 12 & 7 & 4\\ 24 & 16 & 7 \end{bmatrix}$$

of Example CPMS3 [961], Example EMS3 [962] and Example ESMS3 [966]. F's eigenvalues and eigenspaces are

$$\lambda = 3 \qquad \qquad \mathcal{E}_F(3) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = -1 \qquad \qquad \mathcal{E}_F(-1) = \left\langle \left\{ \begin{bmatrix} -\frac{2}{3} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{3} \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$

Define the matrix S to be the 3×3 matrix whose columns are the three basis vectors in the eigenspaces for F,

$$S = \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Check that S is nonsingular (row-reduces to the identity matrix, Theorem NMRRI [176] or has a nonzero determinant, Theorem SMZD [935]). Then the three columns of S are a linearly independent set (Theorem NMLIC [330]). By Theorem DC [1039] we now know that F is diagonalizable. Furthermore, the construction in the proof of Theorem DC [1039] tells us that if we apply the matrix S to F in a similarity transformation, the result will be a diagonal matrix with the eigenvalues of F on the diagonal. The eigenvalues appear on the diagonal of the matrix in the same order as the eigenvectors appear in S. So,

$$S^{-1}FS = \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 6 & 4 & 2 \\ -3 & -1 & -1 \\ -6 & -4 & -1 \end{bmatrix} \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Note that the above computations can be viewed two ways. The proof of Theorem DC [1039] tells us that the four matrices $(F, S, F^{-1} \text{ and the diagonal matrix})$ will interact the way we have written the equation. Or as an example, we can actually *perform* the computations to verify what the theorem predicts.

The dimension of an eigenspace can be no larger than the algebraic multiplicity of the

eigenvalue by Theorem ME [1016]. When every eigenvalue's eigenspace is this large, then we can diagonalize the matrix, and only then. Three examples we have seen so far in this section, Example SMS5 [1030], Example DAB [1037] and Example DMS3 [1042], illustrate the diagonalization of a matrix, with varying degrees of detail about just how the diagonalization is achieved. However, in each case, you can verify that the geometric and algebraic multiplicities are equal for every eigenvalue. This is the substance of the next theorem.

Theorem DMFE

Diagonalizable Matrices have Full Eigenspaces

Suppose A is a square matrix. Then A is diagonalizable if and only if $\gamma_A(\lambda) = \alpha_A(\lambda)$ for every eigenvalue λ of A.

Proof Suppose A has size n and k distinct eigenvalues, $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$. Let $S_i = \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3}, \ldots, \mathbf{x}_{i\gamma_A(\lambda_i)}\}$, denote a basis for the eigenspace of $\lambda_i, \mathcal{E}_A(\lambda_i)$, for $1 \leq i \leq k$. Then

 $S = S_1 \cup S_2 \cup S_3 \cup \cdots \cup S_k$

is a set of eigenvectors for A. A vector cannot be an eigenvector for two different eigenvalues (see Exercise EE.T20 [988]) so $S_i \cap S_j = \emptyset$ whenever $i \neq j$. In other words, S is a disjoint union of S_i , $1 \leq i \leq k$.

 (\Leftarrow) The size of S is

$$|S| = \sum_{i=1}^{k} \gamma_A(\lambda_i) \qquad \qquad S \text{ disjoint union of } S_i$$

$$= \sum_{i=1}^{k} \alpha_A(\lambda_i)$$
 Hypothesis
= n Theorem NEM [1015]

We next show that S is a linearly independent set. So we will begin with a relation of linear dependence on S, using doubly-subscripted scalars and eigenvectors,

$$\mathbf{0} = (a_{11}\mathbf{x}_{11} + a_{12}\mathbf{x}_{12} + \dots + a_{1\gamma_A(\lambda_1)}\mathbf{x}_{1\gamma_A(\lambda_1)}) + (a_{21}\mathbf{x}_{21} + a_{22}\mathbf{x}_{22} + \dots + a_{2\gamma_A(\lambda_2)}\mathbf{x}_{2\gamma_A(\lambda_2)}) + \dots + (a_{k1}\mathbf{x}_{k1} + a_{k2}\mathbf{x}_{k2} + \dots + a_{k\gamma_A(\lambda_k)}\mathbf{x}_{k\gamma_A(\lambda_k)})$$

Define the vectors \mathbf{y}_i , $1 \leq i \leq k$ by

$$\mathbf{y}_{1} = (a_{11}\mathbf{x}_{11} + a_{12}\mathbf{x}_{12} + a_{13}\mathbf{x}_{13} + \dots + a_{\gamma_{A}(1\lambda_{1})}\mathbf{x}_{1\gamma_{A}(\lambda_{1})})$$

$$\mathbf{y}_{2} = (a_{21}\mathbf{x}_{21} + a_{22}\mathbf{x}_{22} + a_{23}\mathbf{x}_{23} + \dots + a_{\gamma_{A}(2\lambda_{2})}\mathbf{x}_{2\gamma_{A}(\lambda_{2})})$$

$$\mathbf{y}_{3} = (a_{31}\mathbf{x}_{31} + a_{32}\mathbf{x}_{32} + a_{33}\mathbf{x}_{33} + \dots + a_{\gamma_{A}(3\lambda_{3})}\mathbf{x}_{3\gamma_{A}(\lambda_{3})})$$

$$\vdots$$

$$\mathbf{y}_{k} = (a_{k1}\mathbf{x}_{k1} + a_{k2}\mathbf{x}_{k2} + a_{k3}\mathbf{x}_{k3} + \dots + a_{\gamma_{A}(k\lambda_{k})}\mathbf{x}_{k\gamma_{A}(\lambda_{k})})$$

Then the relation of linear dependence becomes

$$\mathbf{0} = \mathbf{y}_1 + \mathbf{y}_2 + \mathbf{y}_3 + \dots + \mathbf{y}_k$$

Since the eigenspace $\mathcal{E}_A(\lambda_i)$ is closed under vector addition and scalar multiplication, $\mathbf{y}_i \in \mathcal{E}_A(\lambda_i)$, $1 \leq i \leq k$. Thus, for each *i*, the vector \mathbf{y}_i is an eigenvector of *A* for λ_i , or is the zero vector. Recall that sets of eigenvectors whose eigenvalues are distinct form a linearly independent set by Theorem EDELI [1001]. Should any (or some) \mathbf{y}_i be nonzero, the previous equation would provide a nontrivial relation of linear dependence on a set of eigenvectors with distinct eigenvalues, contradicting Theorem EDELI [1001]. Thus $\mathbf{y}_i = \mathbf{0}$, $1 \leq i \leq k$.

Each of the k equations, $\mathbf{y}_i = \mathbf{0}$ is a relation of linear dependence on the corresponding set S_i , a set of basis vectors for the eigenspace $\mathcal{E}_A(\lambda_i)$, which is therefore linearly independent. From these relations of linear dependence on linearly independent sets we conclude that the scalars are all zero, more precisely, $a_{ij} = 0$, $1 \leq j \leq \gamma_A(\lambda_i)$ for $1 \leq i \leq k$. This establishes that our original relation of linear dependence on S has only the trivial relation of linear dependence, and hence S is a linearly independent set.

We have determined that S is a set of n linearly independent eigenvectors for A, and so by Theorem DC [1039] is diagonalizable.

(⇒) Now we assume that A is diagonalizable. Aiming for a contradiction (Technique CD [1623]), suppose that there is at least one eigenvalue, say λ_t , such that $\gamma_A(\lambda_t) \neq \alpha_A(\lambda_t)$. By Theorem ME [1016] we must have $\gamma_A(\lambda_t) < \alpha_A(\lambda_t)$, and $\gamma_A(\lambda_i) \leq \alpha_A(\lambda_i)$ for $1 \leq i \leq k$, $i \neq t$.

Since A is diagonalizable, Theorem DC [1039] guarantees a set of n linearly independent vectors, all of which are eigenvectors of A. Let n_i denote the number of eigenvectors in S that are eigenvectors for λ_i , and recall that a vector cannot be an eigenvector for two different eigenvalues (Exercise EE.T20 [988]). S is a linearly independent set, so the subset S_i containing the n_i eigenvectors for λ_i must also be linearly independent. Because the

eigenspace $\mathcal{E}_A(\lambda_i)$ has dimension $\gamma_A(\lambda_i)$ and S_i is a linearly independent subset in $\mathcal{E}_A(\lambda_i)$, Theorem G [852] tells us that $n_i \leq \gamma_A(\lambda_i)$, for $1 \leq i \leq k$. Putting all these facts together gives,

$$n = n_{1} + n_{2} + n_{3} + \dots + n_{t} + \dots + n_{k}$$

$$\leq \gamma_{A}(\lambda_{1}) + \gamma_{A}(\lambda_{2}) + \gamma_{A}(\lambda_{3}) + \dots + \gamma_{A}(\lambda_{t}) + \dots + \gamma_{A}(\lambda_{k})$$
Theorem G [852]
$$< \alpha_{A}(\lambda_{1}) + \alpha_{A}(\lambda_{2}) + \alpha_{A}(\lambda_{3}) + \dots + \alpha_{A}(\lambda_{t}) + \dots + \alpha_{A}(\lambda_{k})$$
Theorem ME [1016]
$$= n$$
Theorem NEM [1015]

This is a contradiction (we can't have n < n!) and so our assumption that some eigenspace had less than full dimension was false.

Example SEE [946], Example CAEHW [955], Example ESMS3 [966], Example ESMS4 [970], Example DEMS5 [979], Archetype B [1652], Archetype F [1686], Archetype K [1734] and Archetype L [1742] are all examples of matrices that are diagonalizable and that illustrate Theorem DMFE [1044]. While we have provided many examples of matrices that are diagonalizable, especially among the archetypes, there are many matrices that are not diagonalizable. Here's one now.

Example NDMS4 A non-diagonalizable matrix of size 4 In Example EMMS4 [969] the matrix

$$B = \begin{bmatrix} -2 & 1 & -2 & -4\\ 12 & 1 & 4 & 9\\ 6 & 5 & -2 & -4\\ 3 & -4 & 5 & 10 \end{bmatrix}$$

was determined to have characteristic polynomial

$$p_B(x) = (x-1)(x-2)^3$$

and an eigenspace for $\lambda = 2$ of

$$\mathcal{E}_B(2) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \\ 1 \end{bmatrix} \right\} \right\rangle$$

So the geometric multiplicity of $\lambda = 2$ is $\gamma_B(2) = 1$, while the algebraic multiplicity is $\alpha_B(2) = 3$. By Theorem DMFE [1044], the matrix *B* is not diagonalizable.

Archetype A [1642] is the lone archetype with a square matrix that is not diagonalizable, as the algebraic and geometric multiplicities of the eigenvalue $\lambda = 0$ differ. Example HMEM5 [973] is another example of a matrix that cannot be diagonalized due to the difference between the geometric and algebraic multiplicities of $\lambda = 2$, as is Example CEMS6 [974] which has two complex eigenvalues, each with differing multiplicities. Likewise, Example EMMS4 [969] has an eigenvalue with different algebraic and geometric multiplicities and so cannot be diagonalized.

Theorem DED

Distinct Eigenvalues implies Diagonalizable

Suppose A is a square matrix of size n with n distinct eigenvalues. Then A is diagonalizable.

Proof Let $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ denote the *n* distinct eigenvalues of *A*. Then by Theorem NEM [1015] we have $n = \sum_{i=1}^{n} \alpha_A(\lambda_i)$, which implies that $\alpha_A(\lambda_i) = 1, 1 \le i \le n$. From Theorem ME [1016] it follows that $\gamma_A(\lambda_i) = 1, 1 \le i \le n$. So $\gamma_A(\lambda_i) = \alpha_A(\lambda_i), 1 \le i \le n$ and Theorem DMFE [1044] says *A* is diagonalizable.

Example DEHD Distinct eigenvalues, hence diagonalizable

In Example DEMS5 [979] the matrix

$$H = \begin{bmatrix} 15 & 18 & -8 & 6 & -5\\ 5 & 3 & 1 & -1 & -3\\ 0 & -4 & 5 & -4 & -2\\ -43 & -46 & 17 & -14 & 15\\ 26 & 30 & -12 & 8 & -10 \end{bmatrix}$$

has characteristic polynomial

$$p_H(x) = x(x-2)(x-1)(x+1)(x+3)$$

and so is a 5×5 matrix with 5 distinct eigenvalues. By Theorem DED [1048] we know H must be diagonalizable. But just for practice, we exhibit the diagonalization itself. The matrix S contains eigenvectors of H as columns, one from each eigenspace, guaranteeing linear independent columns and thus the nonsingularity of S. The diagonal matrix has the eigenvalues of H in the same order that their respective eigenvectors appear as the columns of S. Notice that we are using the versions of the eigenvectors from Example DEMS5 [979] that have integer entries.

$S^{-1}HS$

	$\lceil 2 \rangle$	1	-1	1	1]	$^{-1}$	15 1	8 -8	8 6	-5] [2	1	-1	1	1]
	-1	0	2	0	-1		5	3 1	-1	-3	-	1 0	2	0	-1
=	-2	0	2	-1	-2		0 –	-4 5	-4	-2	-	2 0	2	-1	-2
	-4	-1	0	-2	-1	-	-43 -	46 17	7 - 14	4 15	-	4 -1	0	-2	-1
	2	2	1	2	1	L	26 3	50 -1	2 8	-10) [2	2	1	2	1
	$\left[-3\right]$	-3	1	-1	1]	[15	18	-8	6	-5]	$\begin{bmatrix} 2 \end{bmatrix}$	1	-1	1 1	1]
	-1	-2	1	0	1	5	3	1	-1	-3	-1	0	2	0 –	-1
=	-5	-4	1	-1	2	0	-4	5	-4	-2	-2	0	2 -	-1 –	-2
	10	10	-3	2	-4	-4	3 - 46	17	-14	15	-4	-1	0 –	-2 -	-1
	$\lfloor -7 \rfloor$	-6	1	-1	3	2 6	30	-12	8	-10	$\lfloor 2$	2	1 1	2	1

 \boxtimes

Archetype B [1652] is another example of a matrix that has as many distinct eigenvalues as its size, and is hence diagonalizable by Theorem DED [1048].

Powers of a diagonal matrix are easy to compute, and when a matrix is diagonalizable, it is almost as easy. We could state a theorem here perhaps, but we will settle instead for an example that makes the point just as well.

Example HPDM High power of a diagonalizable matrix

Suppose that

$$A = \begin{bmatrix} 19 & 0 & 6 & 13 \\ -33 & -1 & -9 & -21 \\ 21 & -4 & 12 & 21 \\ -36 & 2 & -14 & -28 \end{bmatrix}$$

and we wish to compute A^{20} . Normally this would require 19 matrix multiplications, but since A is diagonalizable, we can simplify the computations substantially. First, we diago-

nalize A. With

$$S = \begin{bmatrix} 1 & -1 & 2 & -1 \\ -2 & 3 & -3 & 3 \\ 1 & 1 & 3 & 3 \\ -2 & 1 & -4 & 0 \end{bmatrix}$$

we find

$$D = S^{-1}AS = \begin{bmatrix} -6 & 1 & -3 & -6 \\ 0 & 2 & -2 & -3 \\ 3 & 0 & 1 & 2 \\ -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 19 & 0 & 6 & 13 \\ -33 & -1 & -9 & -21 \\ 21 & -4 & 12 & 21 \\ -36 & 2 & -14 & -28 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 & -1 \\ -2 & 3 & -3 & 3 \\ 1 & 1 & 3 & 3 \\ -2 & 1 & -4 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now we find an alternate expression for A^{20} ,

$$A^{20} = AAA...A$$

= $I_nAI_nAI_nAI_n...I_nAI_n$
= $(SS^{-1}) A (SS^{-1}) A (SS^{-1}) A (SS^{-1}) ... (SS^{-1}) A (SS^{-1})$
= $S (S^{-1}AS) (S^{-1}AS) (S^{-1}AS) ... (S^{-1}AS) S^{-1}$
= $SDDD...DS^{-1}$

$$= SD^{20}S^{-1}$$

and since D is a diagonal matrix, powers are much easier to compute,

$$= S \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{20} S^{-1}$$

$$= S \begin{bmatrix} (-1)^{20} & 0 & 0 & 0 \\ 0 & (0)^{20} & 0 & 0 \\ 0 & 0 & (2)^{20} & 0 \\ 0 & 0 & 0 & (1)^{20} \end{bmatrix} S^{-1}$$

$$= \begin{bmatrix} 1 & -1 & 2 & -1 \\ -2 & 3 & -3 & 3 \\ 1 & 1 & 3 & 3 \\ -2 & 1 & -4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1048576 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -6 & 1 & -3 & -6 \\ 0 & 2 & -2 & -3 \\ 3 & 0 & 1 & 2 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 6291451 & 2 & 2097148 & 4194297 \\ -9437175 & -5 & -3145719 & -6291441 \\ 9437175 & -2 & 3145728 & 6291453 \\ -12582900 & -2 & -4194298 & -8388596 \end{bmatrix}$$

Notice how we effectively replaced the twentieth power of A by the twentieth power of D, and how a high power of a diagonal matrix is just a collection of powers of scalars on the diagonal.

The price we pay for this simplification is the need to diagonalize the matrix (by computing eigenvalues and eigenvectors) and finding the inverse of the matrix of eigenvectors. And we still need to do two matrix products. But the higher the power, the greater the savings. \boxtimes

Subsection FS Fibonacci Sequences

Example FSCF Fibonacci sequence, closed form The Fibonacci sequence is a sequence of integers defined recursively by

$$a_0 = 0$$
 $a_1 = 1$ $a_{n+1} = a_n + a_{n-1}, n \ge 1$

So the initial portion of the sequence is 0, 1, 1, 2, 3, 5, 8, 13, 21, In this subsection we will illustrate an application of eigenvalues and diagonalization through the determination of a closed-form expression for an arbitrary term of this sequence.

To begin, verify that for any $n \ge 1$ the recursive statement above establishes the truth of the statement

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix}$$

Let A denote this 2×2 matrix. Through repeated applications of the statement above we have

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix} = A^2 \begin{bmatrix} a_{n-2} \\ a_{n-1} \end{bmatrix} = A^3 \begin{bmatrix} a_{n-3} \\ a_{n-2} \end{bmatrix} = \dots = A^n \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

In preparation for working with this high power of A, not unlike in Example HPDM [1051], we will diagonalize A. The characteristic polynomial of A is $p_A(x) = x^2 - x - 1$, with roots (the eigenvalues of A by Theorem EMRCP [962])

$$\rho = \frac{1+\sqrt{5}}{2} \qquad \qquad \delta = \frac{1-\sqrt{5}}{2}$$

With two distinct eigenvalues, Theorem DED [1048] implies that A is diagonalizable. It will be easier to compute with these eigenvalues once you confirm the following properties (all but the last can be derived from the fact that ρ and δ are roots of the characteristic polynomial, in a factored or unfactored form)

$$\rho + \delta = 1$$
 $\rho \delta = -1$ $1 + \rho = \rho^2$ $1 + \delta = \delta^2$ $\rho - \delta = \sqrt{5}$

Then eigenvectors of A (for ρ and δ , respectively) are

$$\begin{bmatrix} 1\\ \rho \end{bmatrix} \qquad \qquad \begin{bmatrix} 1\\ \delta \end{bmatrix}$$

which can be easily confirmed, as we demonstrate for the eigenvector for ρ ,

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \rho \end{bmatrix} = \begin{bmatrix} \rho \\ 1+\rho \end{bmatrix} = \begin{bmatrix} \rho \\ \rho^2 \end{bmatrix} = \rho \begin{bmatrix} 1 \\ \rho \end{bmatrix}$$

From the proof of Theorem DC [1039] we know A can be diagonalized by a matrix S with these eigenvectors as columns, giving $D = S^{-1}AS$. We list S, S^{-1} and the diagonal matrix D,

$$S = \begin{bmatrix} 1 & 1 \\ \rho & \delta \end{bmatrix} \qquad \qquad S^{-1} = \frac{1}{\rho - \delta} \begin{bmatrix} -\delta & 1 \\ \rho & -1 \end{bmatrix} \qquad \qquad D = \begin{bmatrix} \rho & 0 \\ 0 & \delta \end{bmatrix}$$

OK, we have everything in place now. The main step in the following is to replace A by SDS^{-1} . Here we go,

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A^n \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$
$$= (SDS^{-1})^n \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$
$$= SDS^{-1}SDS^{-1}SDS^{-1} \cdots SDS^{-1} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$
$$= SDDD \cdots DS^{-1} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

$$= SD^{n}S^{-1} \begin{bmatrix} a_{0} \\ a_{1} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 \\ \rho & \delta \end{bmatrix} \begin{bmatrix} \rho & 0 \\ 0 & \delta \end{bmatrix}^{n} \frac{1}{\rho - \delta} \begin{bmatrix} -\delta & 1 \\ \rho & -1 \end{bmatrix} \begin{bmatrix} a_{0} \\ a_{1} \end{bmatrix}$$

$$= \frac{1}{\rho - \delta} \begin{bmatrix} 1 & 1 \\ \rho & \delta \end{bmatrix} \begin{bmatrix} \rho^{n} & 0 \\ 0 & \delta^{n} \end{bmatrix} \begin{bmatrix} -\delta & 1 \\ \rho & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{\rho - \delta} \begin{bmatrix} 1 & 1 \\ \rho & \delta \end{bmatrix} \begin{bmatrix} \rho^{n} & 0 \\ 0 & \delta^{n} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$= \frac{1}{\rho - \delta} \begin{bmatrix} 1 & 1 \\ \rho & \delta \end{bmatrix} \begin{bmatrix} \rho^{n} \\ -\delta^{n} \end{bmatrix}$$

$$= \frac{1}{\rho - \delta} \begin{bmatrix} \rho^{n} - \delta^{n} \\ \rho^{n+1} - \delta^{n+1} \end{bmatrix}$$

Performing the scalar multiplication and equating the first entries of the two vectors, we arrive at the closed form expression

$$a_n = \frac{1}{\rho - \delta} \left(\rho^n - \delta^n \right)$$
$$= \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right)$$
$$= \frac{1}{2^n \sqrt{5}} \left(\left(1 + \sqrt{5} \right)^n - \left(1 - \sqrt{5} \right)^n \right)$$

Notice that it does not matter whether we use the equality of the first or second entries of the vectors, we will arrive at the same formula, once in terms of n and again in terms of n + 1. Also, our definition clearly describes a sequence that will only contain integers, yet the presence of the irrational number $\sqrt{5}$ might make us suspicious. But no, our expression for a^n will always yield an integer!

The Fibonacci sequence, and generalizations of it, have been extensively studied (Fibonacci lived in the 12th and 13th centuries). There are many ways to derive the closed-form expression we just found, and our approach may not be the most efficient route. But it is a nice demonstration of how diagonalization can be used to solve a problem outside the field of linear algebra. \boxtimes

We close this section with a comment about an important upcoming theorem that we prove in Chapter R [1260]. A consequence of Theorem OD [1429] is that every Hermitian matrix (Definition HM [486]) is diagonalizable (Definition DZM [1037]), and the similarity transformation that accomplishes the diagonalization uses a unitary matrix (Definition UM [544]). This means that for every Hermitian matrix of size n there is a basis of \mathbb{C}^n that is composed entirely of eigenvectors for the matrix and also forms an orthonormal set (Definition ONS [418]). Notice that for matrices with only real entries, we only need the hypothesis that the matrix is symmetric (Definition SYM [434]) to reach this conclusion (Example ESMS4 [970]). Can you imagine a prettier basis for use with a matrix? I can't.

These results in Section OD [1414] explain much of our recurring interest in orthogonality, and make the section a high point in your study of linear algebra. A precise statement of this diagonalization result applies to a slightly broader class of matrices, known as "normal"

matrices (Definition NRML [1427]), which are matrices that commute with their adjoints. With this expanded category of matrices, the result becomes an equivalence (Technique E [1619]). See Theorem OD [1429] and Theorem OBNM [1434] in Section OD [1414] for all the details.

Subsection READ Reading Questions

- 1. What is an equivalence relation?
- 2. State a condition that is equivalent to a matrix being diagonalizable, but is not the definition.
- 3. Find a diagonal matrix similar to

$$A = \begin{bmatrix} -5 & 8\\ -4 & 7 \end{bmatrix}$$

Subsection EXC Exercises

C20 Consider the matrix A below. First, show that A is diagonalizable by computing the geometric multiplicities of the eigenvalues and quoting the relevant theorem. Second, find a diagonal matrix D and a nonsingular matrix S so that $S^{-1}AS = D$. (See Exercise EE.C20 [985] for some of the necessary computations.)

$$A = \begin{bmatrix} 18 & -15 & 33 & -15 \\ -4 & 8 & -6 & 6 \\ -9 & 9 & -16 & 9 \\ 5 & -6 & 9 & -4 \end{bmatrix}$$

Contributed by Robert Beezer Solution [1063]

C21 Determine if the matrix A below is diagonalizable. If the matrix is diagonalizable, then find a diagonal matrix D that is similar to A, and provide the invertible matrix S that performs the similarity transformation. You should use your calculator to find the eigenvalues of the matrix, but try only using the row-reducing function of your calculator to

assist with finding eigenvectors.

$$A = \begin{bmatrix} 1 & 9 & 9 & 24 \\ -3 & -27 & -29 & -68 \\ 1 & 11 & 13 & 26 \\ 1 & 7 & 7 & 18 \end{bmatrix}$$

Contributed by Robert Beezer Solution [1064]

C22 Consider the matrix A below. Find the eigenvalues of A using a calculator and use these to construct the characteristic polynomial of A, $p_A(x)$. State the algebraic multiplicity of each eigenvalue. Find all of the eigenspaces for A by computing expressions for null spaces, only using your calculator to row-reduce matrices. State the geometric multiplicity of each eigenvalue. Is A diagonalizable? If not, explain why. If so, find a diagonal matrix D that is similar to A.

$$A = \begin{bmatrix} 19 & 25 & 30 & 5 \\ -23 & -30 & -35 & -5 \\ 7 & 9 & 10 & 1 \\ -3 & -4 & -5 & -1 \end{bmatrix}$$

Contributed by Robert Beezer Solution [1066]

T15 Suppose that A and B are similar matrices. Prove that A^3 and B^3 are similar matrices. Generalize.

Contributed by Robert Beezer Solution [1068]

T16 Suppose that A and B are similar matrices, with A nonsingular. Prove that B is nonsingular, and that A^{-1} is similar to B^{-1} . Contributed by Robert Beezer Solution [1069]

T17 Suppose that B is a nonsingular matrix. Prove that AB is similar to BA. Contributed by Robert Beezer Solution [1069]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [1060]

Using a calculator, we find that A has three distinct eigenvalues, $\lambda = 3, 2, -1$, with $\lambda = 2$ having algebraic multiplicity two, $\alpha_A(2) = 2$. The eigenvalues $\lambda = 3, -1$ have algebraic multiplicity one, and so by Theorem ME [1016] we can conclude that their geometric multiplicities are one as well. Together with the computation of the geometric multiplicity of $\lambda = 2$ from Exercise EE.C20 [985], we know

 $\gamma_A(3) = \alpha_A(3) = 1$ $\gamma_A(2) = \alpha_A(2) = 2$ $\gamma_A(-1) = \alpha_A(-1) = 1$

This satisfies the hypotheses of Theorem DMFE [1044], and so we can conclude that A is diagonalizable.

A calculator will give us four eigenvectors of A, the two for $\lambda = 2$ being linearly independent presumably. Or, by hand, we could find basis vectors for the three eigenspaces. For $\lambda = 3, -1$ the eigenspaces have dimension one, and so any eigenvector for these eigenvalues will be multiples of the ones we use below. For $\lambda = 2$ there are many different bases for the eigenspace, so your answer could vary. Our eigenvectors are the basis vectors we would have obtained if we had actually constructed a basis in Exercise EE.C20 [985] rather than just computing the dimension.

By the construction in the proof of Theorem DC [1039], the required matrix S has

columns that are four linearly independent eigenvectors of A and the diagonal matrix has the eigenvalues on the diagonal (in the same order as the eigenvectors in S). Here are the pieces, "doing" the diagonalization,

$$\begin{bmatrix} -1 & 0 & -3 & 6 \\ -2 & -1 & -1 & 0 \\ 0 & 0 & 1 & -3 \\ 1 & 1 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 18 & -15 & 33 & -15 \\ -4 & 8 & -6 & 6 \\ -9 & 9 & -16 & 9 \\ 5 & -6 & 9 & -4 \end{bmatrix} \begin{bmatrix} -1 & 0 & -3 & 6 \\ -2 & -1 & -1 & 0 \\ 0 & 0 & 1 & -3 \\ 1 & 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

C21 Contributed by Robert Beezer Statement [1060] A calculator will provide the eigenvalues $\lambda = 2, 2, 1, 0$, so we can reconstruct the characteristic polynomial as

$$p_A(x) = (x-2)^2(x-1)x$$

so the algebraic multiplicities of the eigenvalues are

$$\alpha_A(2) = 2 \qquad \qquad \alpha_A(1) = 1 \qquad \qquad \alpha_A(0) = 1$$

Now compute eigenspaces by hand, obtaining null spaces for each of the three eigenvalues by constructing the correct singular matrix (Theorem EMNS [965]),

$$A - 2I_4 = \begin{bmatrix} -1 & 9 & 9 & 24 \\ -3 & -29 & -29 & -68 \\ 1 & 11 & 11 & 26 \\ 1 & 7 & 7 & 16 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -\frac{3}{2} \\ 0 & 1 & 1 & \frac{5}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{split} \mathcal{E}_{A}\left(2\right) &= \mathcal{N}(A - 2I_{4}) = \left\langle \left\{ \begin{bmatrix} \frac{3}{2} \\ -\frac{5}{2} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 3 \\ -5 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle \\ A - 1I_{4} &= \begin{bmatrix} 0 & 9 & 9 & 24 \\ -3 & -28 & -29 & -68 \\ 1 & 11 & 12 & 26 \\ 1 & 7 & 7 & 17 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -\frac{5}{3} \\ 0 & 1 & 0 & \frac{13}{3} \\ 0 & 0 & 1 & -\frac{5}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathcal{E}_{A}\left(1\right) &= \mathcal{N}(A - I_{4}) = \left\langle \left\{ \begin{bmatrix} \frac{5}{3} \\ -\frac{13}{3} \\ \frac{5}{3} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 5 \\ -13 \\ 5 \\ 3 \end{bmatrix} \right\} \right\rangle \\ A - 0I_{4} &= \begin{bmatrix} 1 & 9 & 9 & 24 \\ -3 & -27 & -29 & -68 \\ 1 & 11 & 13 & 26 \\ 1 & 7 & 7 & 18 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\ \mathcal{E}_{A}\left(0\right) &= \mathcal{N}(A - I_{4}) = \left\langle \left\{ \begin{bmatrix} 3 \\ -5 \\ 2 \\ 1 \end{bmatrix} \right\} \right\rangle \end{split}$$

From this we can compute the dimensions of the eigenspaces to obtain the geometric multi-

plicities,

$$\gamma_A(2) = 2 \qquad \qquad \gamma_A(1) = 1 \qquad \qquad \gamma_A(0) = 1$$

For each eigenvalue, the algebraic and geometric multiplicities are equal and so by Theorem DMFE [1044] we now know that A is diagonalizable. The construction in Theorem DC [1039] suggests we form a matrix whose columns are eigenvectors of A

$$S = \begin{bmatrix} 3 & 0 & 5 & 3 \\ -5 & -1 & -13 & -5 \\ 0 & 1 & 5 & 2 \\ 2 & 0 & 3 & 1 \end{bmatrix}$$

Since det $(S) = -1 \neq 0$, we know that S is nonsingular (Theorem SMZD [935]), so the columns of S are a set of 4 linearly independent eigenvectors of A. By the proof of Theorem SMZD [935] we know

$$S^{-1}AS = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

a diagonal matrix with the eigenvalues of A along the diagonal, in the same order as the associated eigenvectors appear as columns of S.

C22 Contributed by Robert Beezer Statement [1061] A calculator will report $\lambda = 0$ as an eigenvalue of algebraic multiplicity of 2, and $\lambda = -1$ as an eigenvalue of algebraic multiplicity 2 as well. Since eigenvalues are roots of the characteristic polynomial (Theorem EMRCP [962]) we have the factored version

$$p_A(x) = (x-0)^2(x-(-1))^2 = x^2(x^2+2x+1) = x^4+2x^3+x^2$$

The eigenspaces are then

$$\lambda = 0$$

$$A - (0)I_4 = \begin{bmatrix} 19 & 25 & 30 & 5 \\ -23 & -30 & -35 & -5 \\ 7 & 9 & 10 & 1 \\ -3 & -4 & -5 & -1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -5 & -5 \\ 0 & 1 & 5 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_A(0) = \mathcal{N}(C - (0)I_4) = \left\langle \left\{ \begin{bmatrix} 5 \\ -5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 5 \\ -4 \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -1$$

$$A - (-1)I_4 = \begin{bmatrix} 20 & 25 & 30 & 5\\ -23 & -29 & -35 & -5\\ 7 & 9 & 11 & 1\\ -3 & -4 & -5 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & 4\\ 0 & 1 & 2 & -3\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_{A}(-1) = \mathcal{N}(C - (-1)I_{4}) = \left\langle \left\{ \begin{bmatrix} 1\\-2\\1\\0 \end{bmatrix}, \begin{bmatrix} -4\\3\\0\\1 \end{bmatrix} \right\} \right\rangle$$

Each eigenspace above is described by a spanning set obtained through an application of Theorem BNS [334] and so is a basis for the eigenspace. In each case the dimension, and therefore the geometric multiplicity, is 2.

For each of the two eigenvalues, the algebraic and geometric multiplicities are equal. Theorem DMFE [1044] says that in this situation the matrix is diagonalizable. We know from Theorem DC [1039] that when we diagonalize A the diagonal matrix will have the eigenvalues of A on the diagonal (in some order). So we can claim that

T15 Contributed by Robert Beezer Statement [1061] By Definition SIM [1029] we know that there is a nonsingular matrix S so that $A = S^{-1}BS$. Then

$$A^{3} = (S^{-1}BS)^{3}$$

= (S^{-1}BS)(S^{-1}BS)(S^{-1}BS)

$$= S^{-1}B(SS^{-1})B(SS^{-1})BS$$
$$= S^{-1}B(I_3)B(I_3)BS$$
$$= S^{-1}BBBS$$
$$= S^{-1}B^3S$$

Theorem MMA [478] Definition MI [506] Theorem MMIM [475]

This equation says that A^3 is similar to B^3 (via the matrix S).

More generally, if A is similar to B, and m is a non-negative integer, then A^m is similar to B^m . This can be proved using induction (Technique I [1630]).

T16 Contributed by Steve Canfield Statement [1062] A being similar to B means that there exists an S such that $A = S^{-1}BS$. So, $B = SAS^{-1}$ and because S, A, and S^{-1} are nonsingular, by Theorem NPNT [538], B is nonsingular.

 $A^{-1} = (S^{-1}BS)^{-1}$ Definition SIM [1029] = $S^{-1}B^{-1}(S^{-1})^{-1}$ Theorem SS [520] = $S^{-1}B^{-1}S$ Theorem MIMI [521]

Then by Definition SIM [1029], A^{-1} is similar to B^{-1} .

T17 Contributed by Robert Beezer Statement [1062] The nonsingular (invertible) matrix B will provide the desired similarity transformation,

$B^{-1}(BA) B = \left(B^{-1}B\right)(AB)$	Theorem MMA [478]
$=I_nAB$	Definition MI [506]
= AB	Theorem MMIM [475]

Chapter LT Linear Transformations

In the next linear algebra course you take, the first lecture might be a reminder about what a vector space is (Definition VS [659]), their ten properties, basic theorems and then some examples. The second lecture would likely be all about linear transformations. While it may seem we have waited a long time to present what must be a central topic, in truth we have already been working with linear transformations for some time.

Functions are important objects in the study of calculus, but have been absent from this course until now (well, not really, it just seems that way). In your study of more advanced mathematics it is nearly impossible to escape the use of functions — they are as fundamental as sets are.

Section LT Linear Transformations

Early in Chapter VS [658] we prefaced the definition of a vector space with the comment that it was "one of the two most important definitions in the entire course." Here comes the other. Any capsule summary of linear algebra would have to describe the subject as the interplay of linear transformations and vector spaces. Here we go.

Subsection LT Linear Transformations

Definition LT Linear Transformation

A linear transformation, $T: U \to V$, is a function that carries elements of the vector space U (called the **domain**) to the vector space V (called the **codomain**), and which has two additional properties

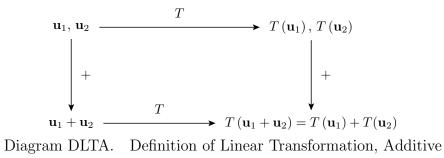
1.
$$T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$$
 for all $\mathbf{u}_1, \mathbf{u}_2 \in U$

2. $T(\alpha \mathbf{u}) = \alpha T(\mathbf{u})$ for all $\mathbf{u} \in U$ and all $\alpha \in \mathbb{C}$

(This definition contains Notation LT.)

The two defining conditions in the definition of a linear transformation should "feel linear," whatever that means. Conversely, these two conditions could be taken as *exactly* what it means *to be* linear. As every vector space property derives from vector addition and scalar multiplication, so too, every property of a linear transformation derives from these two defining properties. While these conditions may be reminiscent of how we test subspaces, they really are quite different, so do not confuse the two.

Here are two diagrams that convey the essence of the two defining properties of a linear transformation. In each case, begin in the upper left-hand corner, and follow the arrows around the rectangle to the lower-right hand corner, taking two different routes and doing the indicated operations labeled on the arrows. There are two results there. For a linear transformation these two expressions are always equal.



 \triangle

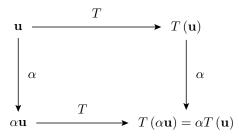


Diagram DLTM. Definition of Linear Transformation, Multiplicative

A couple of words about notation. T is the *name* of the linear transformation, and should be used when we want to discuss the function as a whole. $T(\mathbf{u})$ is how we talk about the output of the function, it is a vector in the vector space V. When we write $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$, the plus sign on the left is the operation of vector addition in the vector space U, since \mathbf{x} and \mathbf{y} are elements of U. The plus sign on the right is the operation of vector addition in the vector space V, since $T(\mathbf{x})$ and $T(\mathbf{y})$ are elements of the vector space V. These two instances of vector addition might be wildly different.

Let's examine several examples and begin to form a catalog of known linear transformations to work with.

Example ALT A linear transformation

Define $T: \mathbb{C}^3 \to \mathbb{C}^2$ by describing the output of the function for a generic input with the

formula

$$T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}2x_1+x_3\\-4x_2\end{bmatrix}$$

and check the two defining properties.

$$T (\mathbf{x} + \mathbf{y}) = T \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right)$$
$$= T \left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{bmatrix} \right)$$
$$= \begin{bmatrix} 2(x_1 + y_1) + (x_3 + y_3) \\ -4(x_2 + y_2) \end{bmatrix}$$
$$= \begin{bmatrix} (2x_1 + x_3) + (2y_1 + y_3) \\ -4x_2 + (-4)y_2 \end{bmatrix}$$
$$= \begin{bmatrix} 2x_1 + x_3 \\ -4x_2 \end{bmatrix} + \begin{bmatrix} 2y_1 + y_3 \\ -4y_2 \end{bmatrix}$$
$$= T \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) + T \left(\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right)$$
$$= T (\mathbf{x}) + T (\mathbf{y})$$

and

$$(\alpha \mathbf{x}) = T \left(\alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right)$$
$$= T \left(\begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{bmatrix} \right)$$
$$= \begin{bmatrix} 2(\alpha x_1) + (\alpha x_3) \\ -4(\alpha x_2) \end{bmatrix}$$
$$= \begin{bmatrix} \alpha(2x_1 + x_3) \\ \alpha(-4x_2) \end{bmatrix}$$
$$= \alpha \begin{bmatrix} 2x_1 + x_3 \\ \alpha(-4x_2) \end{bmatrix}$$
$$= \alpha T \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right)$$
$$= \alpha T (\mathbf{x})$$

So by Definition LT [1072], T is a linear transformation.

T

It can be just as instructive to look at functions that are *not* linear transformations. Since

 \boxtimes

the defining conditions must be true for *all* vectors and scalars, it is enough to find just one situation where the properties fail.

Example NLT Not a linear transformation Define $S: \mathbb{C}^3 \to \mathbb{C}^3$ by

$$S\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}4x_1+2x_2\\0\\x_1+3x_3-2\end{bmatrix}$$

This function "looks" linear, but consider

$$3S\left(\begin{bmatrix}1\\2\\3\end{bmatrix}\right) = 3\begin{bmatrix}8\\0\\8\end{bmatrix} = \begin{bmatrix}24\\0\\24\end{bmatrix}$$

while

$$S\left(3 \begin{bmatrix} 1\\2\\3 \end{bmatrix}\right) = S\left(\begin{bmatrix} 3\\6\\9 \end{bmatrix}\right) = \begin{bmatrix} 24\\0\\28 \end{bmatrix}$$

So the second required property fails for the choice of $\alpha = 3$ and $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and by Definition LT [1072], S is not a linear transformation. It is just about as easy to find an example where

the first defining property fails (try it!). Notice that it is the "-2" in the third component of the definition of S that prevents the function from being a linear transformation.

Example LTPM Linear transformation, polynomials to matrices Define a linear transformation $T: P_3 \rightarrow M_{22}$ by

$$T(a+bx+cx^{2}+dx^{3}) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

We verify the two defining conditions of a linear transformations.

$$T (\mathbf{x} + \mathbf{y}) = T \left((a_1 + b_1 x + c_1 x^2 + d_1 x^3) + (a_2 + b_2 x + c_2 x^2 + d_2 x^3) \right)$$

= $T \left((a_1 + a_2) + (b_1 + b_2) x + (c_1 + c_2) x^2 + (d_1 + d_2) x^3 \right)$
= $\begin{bmatrix} (a_1 + a_2) + (b_1 + b_2) & (a_1 + a_2) - 2(c_1 + c_2) \\ d_1 + d_2 & (b_1 + b_2) - (d_1 + d_2) \end{bmatrix}$
= $\begin{bmatrix} (a_1 + b_1) + (a_2 + b_2) & (a_1 - 2c_1) + (a_2 - 2c_2) \\ d_1 + d_2 & (b_1 - d_1) + (b_2 - d_2) \end{bmatrix}$
= $\begin{bmatrix} a_1 + b_1 & a_1 - 2c_1 \\ d_1 & b_1 - d_1 \end{bmatrix} + \begin{bmatrix} a_2 + b_2 & a_2 - 2c_2 \\ d_2 & b_2 - d_2 \end{bmatrix}$
= $T \left(a_1 + b_1 x + c_1 x^2 + d_1 x^3 \right) + T \left(a_2 + b_2 x + c_2 x^2 + d_2 x^3 \right)$
= $T (\mathbf{x}) + T (\mathbf{y})$

and

$$T (\alpha \mathbf{x}) = T (\alpha (a + bx + cx^{2} + dx^{3}))$$

$$= T ((\alpha a) + (\alpha b)x + (\alpha c)x^{2} + (\alpha d)x^{3})$$

$$= \begin{bmatrix} (\alpha a) + (\alpha b) & (\alpha a) - 2(\alpha c) \\ \alpha d & (\alpha b) - (\alpha d) \end{bmatrix}$$

$$= \begin{bmatrix} \alpha (a + b) & \alpha (a - 2c) \\ \alpha d & \alpha (b - d) \end{bmatrix}$$

$$= \alpha \begin{bmatrix} a + b & a - 2c \\ d & b - d \end{bmatrix}$$

$$= \alpha T (a + bx + cx^{2} + dx^{3})$$

$$= \alpha T (\mathbf{x})$$

So by Definition LT [1072], T is a linear transformation.

Example LTPP Linear transformation, polynomials to polynomials Define a function $S: P_4 \rightarrow P_5$ by

$$S(p(x)) = (x-2)p(x)$$

Then

$$S(p(x) + q(x)) = (x - 2)(p(x) + q(x)) = (x - 2)p(x) + (x - 2)q(x) = S(p(x)) + S(q(x))$$

 \boxtimes

$$S\left(\alpha p(x)\right) = (x-2)(\alpha p(x)) = (x-2)\alpha p(x) = \alpha(x-2)p(x) = \alpha S\left(p(x)\right)$$

So by Definition LT [1072], S is a linear transformation.

Linear transformations have many amazing properties, which we will investigate through the next few sections. However, as a taste of things to come, here is a theorem we can prove now and put to use immediately.

Theorem LTTZZ

Linear Transformations Take Zero to Zero

Suppose $T: U \to V$ is a linear transformation. Then $T(\mathbf{0}) = \mathbf{0}$.

Proof The two zero vectors in the conclusion of the theorem are different. The first is from U while the second is from V. We will subscript the zero vectors in this proof to highlight the distinction. Think about your objects. (This proof is contributed by Mark Shoemaker).

$T\left(0_{U}\right)=T\left(00_{U}\right)$	Theorem ZSSM [676] in U
$=0T\left(0_{U}\right)$	Definition LT $[1072]$
$= 0_V$	Theorem ZSSM [676] in V

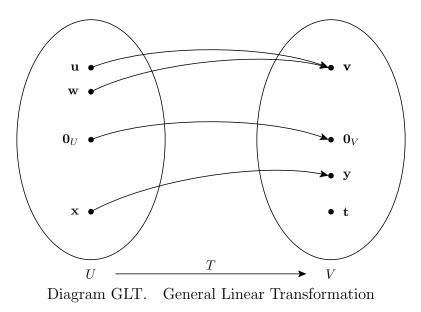
Return to Example NLT [1077] and compute $S\left(\begin{bmatrix}0\\0\\0\end{bmatrix}\right) = \begin{bmatrix}0\\0\\-2\end{bmatrix}$ to quickly see again that *S* is not a linear transformation, while in Example LTPM [1078] compute $S(0 + 0x + 0x^2 + 0x^3) = \begin{bmatrix}0&0\\0&0\end{bmatrix}$ as an example of Theorem LTTZZ [1080] at work.

 \square

 \square

Subsection LTC Linear Transformation Cartoons

Throughout this chapter, and Chapter R [1260], we will include drawings of linear transformations. We will call them "cartoons," not because they are humorous, but because they will only expose a portion of the truth. A Bugs Bunny cartoon might give us some insights on human nature, but the rules of physics and biology are routinely (and grossly) violated. So it will be with our **linear transformation cartoons**. Here is our first, followed by a guide to help you understand how these are meant to describe fundamental truths about linear transformations, while simultaneously violating other truths.



Here we picture a linear transformation $T: U \to V$, where this information will be consistently displayed along the bottom edge. The ovals are meant to represent the vector spaces, in this case U, the domain, on the left and V, the codomain, on the right. Of course, vector spaces are typically infinite sets, so you'll have to imagine that characteristic of these sets. A small dot inside of an oval will represent a vector within that vector space, sometimes with a name, sometimes not (in this case every vector has a name). The sizes of the ovals are meant to be proportional to the dimensions of the vector spaces. However, when we make

no assumptions about the dimensions, we will draw the ovals as the same size, as we have done here (which is not meant to suggest that the dimensions have to be equal).

To convey that the linear transformation associates a certain input with a certain output, we will draw an arrow from the input to the output. So, for example, in this cartoon we suggest that $T(\mathbf{x}) = \mathbf{y}$. Nothing in the definition of a linear transformation prevents two different inputs being sent to the same output and we see this in $T(\mathbf{u}) = \mathbf{v} = T(\mathbf{w})$. Similarly, an output may not have any input being sent its way, as illustrated by no arrow pointing at \mathbf{t} . In this cartoon, we have captured the essence of our one general theorem about linear transformations, Theorem LTTZZ [1080], $T(\mathbf{0}_U) = \mathbf{0}_V$. On occasion we might include this basic fact when it is relevant, at other times maybe not. Note that the definition of a linear transformation requires that it be a function, so every element of the domain should be associated with some element of the codomain. This will be reflected by never having an element of the domain without an arrow originating there.

These cartoons are of course no substitute for careful definitions and proofs, but they can be a handy way to think about the various properties we will be studying.

Subsection MLT Matrices and Linear Transformations

If you give me a matrix, then I can quickly build you a linear transformation. Always. First a motivating example and then the theorem.

Example LTM Linear transformation from a matrix Let

$$A = \begin{bmatrix} 3 & -1 & 8 & 1 \\ 2 & 0 & 5 & -2 \\ 1 & 1 & 3 & -7 \end{bmatrix}$$

and define a function $P\colon \mathbb{C}^4\to \mathbb{C}^3$ by

$$P\left(\mathbf{x}\right) = A\mathbf{x}$$

So we are using an old friend, the matrix-vector product (Definition MVP [461]) as a way to convert a vector with 4 components into a vector with 3 components. Applying Definition MVP [461] allows us to write the defining formula for P in a slightly different form,

$$P(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 3 & -1 & 8 & 1 \\ 2 & 0 & 5 & -2 \\ 1 & 1 & 3 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_1 \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 8 \\ 5 \\ 3 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ -2 \\ -7 \end{bmatrix}$$

So we recognize the action of the function P as using the components of the vector (x_1, x_2, x_3, x_4) as scalars to form the output of P as a linear combination of the four columns of the matrix A, which are all members of \mathbb{C}^3 , so the result is a vector in \mathbb{C}^3 . We can rearrange this expression further, using our definitions of operations in \mathbb{C}^3 (Section VO [198]).

$$P(\mathbf{x}) = A\mathbf{x}$$
 Definition of P

$$= x_{1} \begin{bmatrix} 3\\2\\1 \end{bmatrix} + x_{2} \begin{bmatrix} -1\\0\\1 \end{bmatrix} + x_{3} \begin{bmatrix} 8\\5\\3 \end{bmatrix} + x_{4} \begin{bmatrix} 1\\-2\\-7 \end{bmatrix}$$
Definition MVP [461]
$$= \begin{bmatrix} 3x_{1}\\2x_{1}\\x_{1} \end{bmatrix} + \begin{bmatrix} -x_{2}\\0\\x_{2} \end{bmatrix} + \begin{bmatrix} 8x_{3}\\5x_{3}\\3x_{3} \end{bmatrix} + \begin{bmatrix} x_{4}\\-2x_{4}\\-7x_{4} \end{bmatrix}$$
Definition CVSM [203]
$$= \begin{bmatrix} 3x_{1} - x_{2} + 8x_{3} + x_{4}\\2x_{1} + 5x_{3} - 2x_{4}\\x_{1} + x_{2} + 3x_{3} - 7x_{4} \end{bmatrix}$$
Definition CVA [201]

You might recognize this final expression as being similar in style to some previous examples (Example ALT [1074]) and some linear transformations defined in the archetypes (Archetype M [1750] through Archetype R [1779]). But the expression that says the output of this linear transformation is a linear combination of the columns of A is probably the most powerful way of thinking about examples of this type.

Almost forgot — we should verify that P is indeed a linear transformation. This is easy with two matrix properties from Section MM [460].

$P\left(\mathbf{x}+\mathbf{y}\right) = A\left(\mathbf{x}+\mathbf{y}\right)$	Definition of P
$= A\mathbf{x} + A\mathbf{y}$	Theorem MMDAA [476]
$=P\left(\mathbf{x} ight)+P\left(\mathbf{y} ight)$	Definition of P

and

$P\left(\alpha\mathbf{x}\right) = A\left(\alpha\mathbf{x}\right)$	Definition of P
$= \alpha \left(A \mathbf{x} \right)$	Theorem MMSMM [477]
$= \alpha P\left(\mathbf{x}\right)$	Definition of P

So by Definition LT [1072], P is a linear transformation.

So the multiplication of a vector by a matrix "transforms" the input vector into an output vector, possibly of a different size, by performing a linear combination. And this transformation happens in a "linear" fashion. This "functional" view of the matrix-vector product is the most important shift you can make right now in how you think about linear algebra. Here's the theorem, whose proof is very nearly an exact copy of the verification in the last example.

Theorem MBLT

Matrices Build Linear Transformations

Suppose that A is an $m \times n$ matrix. Define a function $T: \mathbb{C}^n \to \mathbb{C}^m$ by $T(\mathbf{x}) = A\mathbf{x}$. Then T is a linear transformation.

Proof

$$T (\mathbf{x} + \mathbf{y}) = A (\mathbf{x} + \mathbf{y})$$

Definition of T
$$= A\mathbf{x} + A\mathbf{y}$$

Theorem MMDAA [476]
$$= T (\mathbf{x}) + T (\mathbf{y})$$

Definition of T

 \boxtimes

and

$T\left(\alpha\mathbf{x}\right) = A\left(\alpha\mathbf{x}\right)$	Definition of T
$= \alpha \left(A \mathbf{x} \right)$	Theorem MMSMM [477]
$= \alpha T \left(\mathbf{x} \right)$	Definition of T

So by Definition LT [1072], T is a linear transformation.

So Theorem MBLT [1086] gives us a rapid way to construct linear transformations. Grab an $m \times n$ matrix A, define $T(\mathbf{x}) = A\mathbf{x}$ and Theorem MBLT [1086] tells us that T is a linear transformation from \mathbb{C}^n to \mathbb{C}^m , without any further checking.

We can turn Theorem MBLT [1086] around. You give me a linear transformation and I will give you a matrix.

Example MFLT Matrix from a linear transformation Define the function $R: \mathbb{C}^3 \to \mathbb{C}^4$ by

$$R\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}2x_1 - 3x_2 + 4x_3\\x_1 + x_2 + x_3\\-x_1 + 5x_2 - 3x_3\\x_2 - 4x_3\end{bmatrix}$$

You could verify that R is a linear transformation by applying the definition, but we will instead massage the expression defining a typical output until we recognize the form of a

known class of linear transformations.

$$R\left(\begin{bmatrix}x_{1}\\x_{2}\\x_{3}\end{bmatrix}\right) = \begin{bmatrix}2x_{1} - 3x_{2} + 4x_{3}\\x_{1} + x_{2} + x_{3}\\-x_{1} + 5x_{2} - 3x_{3}\\x_{2} - 4x_{3}\end{bmatrix}$$
$$= \begin{bmatrix}2x_{1}\\x_{1}\\-x_{1}\\0\end{bmatrix} + \begin{bmatrix}-3x_{2}\\x_{2}\\5x_{2}\\x_{2}\end{bmatrix} + \begin{bmatrix}4x_{3}\\x_{3}\\-3x_{3}\\-4x_{3}\end{bmatrix}$$
Definition CVA [201]
$$= x_{1}\begin{bmatrix}2\\1\\-1\\0\end{bmatrix} + x_{2}\begin{bmatrix}-3\\1\\5\\1\end{bmatrix} + x_{3}\begin{bmatrix}4\\1\\-3\\-4\end{bmatrix}$$
Definition CVSM [203]
$$= \begin{bmatrix}2 & -3 & 4\\1 & 1 & 1\\-1 & 5 & -3\\0 & 1 & -4\end{bmatrix}\begin{bmatrix}x_{1}\\x_{2}\\x_{3}\end{bmatrix}$$
Definition MVP [461]

So if we define the matrix

$$B = \begin{bmatrix} 2 & -3 & 4 \\ 1 & 1 & 1 \\ -1 & 5 & -3 \\ 0 & 1 & -4 \end{bmatrix}$$

then $R(\mathbf{x}) = B\mathbf{x}$. By Theorem MBLT [1086], we can easily recognize R as a linear transformation since it has the form described in the hypothesis of the theorem.

Example MFLT [1087] was not accident. Consider any one of the archetypes where both the domain and codomain are sets of column vectors (Archetype M [1750] through Archetype R [1779]) and you should be able to mimic the previous example. Here's the theorem, which is notable since it is our first occasion to use the full power of the defining properties of a linear transformation when our hypothesis includes a linear transformation.

Theorem MLTCV

Matrix of a Linear Transformation, Column Vectors

Suppose that $T: \mathbb{C}^n \to \mathbb{C}^m$ is a linear transformation. Then there is an $m \times n$ matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

Proof The conclusion says a certain matrix exists. What better way to prove something exists than to actually build it? So our proof will be constructive (Technique C [1618]), and the procedure that we will use abstractly in the proof can be used concretely in specific examples.

Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \ldots, \mathbf{e}_n$ be the columns of the identity matrix of size n, I_n (Definition SUV [409]). Evaluate the linear transformation T with each of these standard unit vectors as an input, and record the result. In other words, define n vectors in $\mathbb{C}^m, \mathbf{A}_i, 1 \leq i \leq n$ by

$$\mathbf{A}_{i}=T\left(\mathbf{e}_{i}\right)$$

Then package up these vectors as the columns of a matrix

$$A = [\mathbf{A}_1 | \mathbf{A}_2 | \mathbf{A}_3 | \dots | \mathbf{A}_n]$$

Does A have the desired properties? First, A is clearly an $m \times n$ matrix. Then

$$T (\mathbf{x}) = T (I_n \mathbf{x})$$
Theorem MMIM [475]

$$= T ([\mathbf{e}_1 | \mathbf{e}_2 | \mathbf{e}_3 | \dots | \mathbf{e}_n] \mathbf{x})$$
Definition SUV [409]

$$= T ([\mathbf{x}]_1 \mathbf{e}_1 + [\mathbf{x}]_2 \mathbf{e}_2 + [\mathbf{x}]_3 \mathbf{e}_3 + \dots + [\mathbf{x}]_n \mathbf{e}_n)$$
Definition MVP [461]

$$= T ([\mathbf{x}]_1 \mathbf{e}_1) + T ([\mathbf{x}]_2 \mathbf{e}_2) + T ([\mathbf{x}]_3 \mathbf{e}_3) + \dots + T ([\mathbf{x}]_n \mathbf{e}_n)$$
Definition LT [1072]

$$= [\mathbf{x}]_1 T (\mathbf{e}_1) + [\mathbf{x}]_2 T (\mathbf{e}_2) + [\mathbf{x}]_3 T (\mathbf{e}_3) + \dots + [\mathbf{x}]_n T (\mathbf{e}_n)$$
Definition LT [1072]

$$= [\mathbf{x}]_1 \mathbf{A}_1 + [\mathbf{x}]_2 \mathbf{A}_2 + [\mathbf{x}]_3 \mathbf{A}_3 + \dots + [\mathbf{x}]_n \mathbf{A}_n$$
Definition of \mathbf{A}_i

$$= A\mathbf{x}$$
Definition MVP [461]

as desired.

So if we were to restrict our study of linear transformations to those where the domain and codomain are both vector spaces of column vectors (Definition VSCV [198]), every matrix leads to a linear transformation of this type (Theorem MBLT [1086]), while every such linear transformation leads to a matrix (Theorem MLTCV [1089]). So matrices and linear transformations are fundamentally the same. We call the matrix A of Theorem MLTCV [1089] the matrix representation of T.

We have defined linear transformations for more general vector spaces than just \mathbb{C}^m , can we extend this correspondence between linear transformations and matrices to more general linear transformations (more general domains and codomains)? Yes, and this is the main theme of Chapter R [1260]. Stay tuned. For now, let's illustrate Theorem MLTCV [1089] with an example. Example MOLT Matrix of a linear transformation Suppose $S: \mathbb{C}^3 \to \mathbb{C}^4$ is defined by

$$S\left(\begin{bmatrix}x_{1}\\x_{2}\\x_{3}\end{bmatrix}\right) = \begin{bmatrix}3x_{1} - 2x_{2} + 5x_{3}\\x_{1} + x_{2} + x_{3}\\9x_{1} - 2x_{2} + 5x_{3}\\4x_{2}\end{bmatrix}$$

Then

$$\mathbf{C}_{1} = S\left(\mathbf{e_{1}}\right) = S\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}3\\1\\9\\0\end{bmatrix}$$
$$\mathbf{C}_{2} = S\left(\mathbf{e_{2}}\right) = S\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}-2\\1\\-2\\4\end{bmatrix}$$
$$\mathbf{C}_{3} = S\left(\mathbf{e_{3}}\right) = S\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) = \begin{bmatrix}5\\1\\5\\0\end{bmatrix}$$

-

so define

$$C = [C_1|C_2|C_3] = \begin{bmatrix} 3 & -2 & 5\\ 1 & 1 & 1\\ 9 & -2 & 5\\ 0 & 4 & 0 \end{bmatrix}$$

and Theorem MLTCV [1089] guarantees that $S(\mathbf{x}) = C\mathbf{x}$.
As an illuminating exercise, let $\mathbf{z} = \begin{bmatrix} 2\\ -3\\ 3 \end{bmatrix}$ and compute $S(\mathbf{z})$ two different ways. First,
return to the definition of S and evaluate $S(\mathbf{z})$ directly. Then do the matrix-vector product
 $C\mathbf{z}$. In both cases you should obtain the vector $S(\mathbf{z}) = \begin{bmatrix} 27\\ 2\\ 39\\ -12 \end{bmatrix}$.

Subsection LTLC Linear Transformations and Linear Combinations

It is the interaction between linear transformations and linear combinations that lies at the heart of many of the important theorems of linear algebra. The next theorem distills the essence of this. The proof is not deep, the result is hardly startling, but it will be

 \boxtimes

referenced frequently. We have already passed by one occasion to employ it, in the proof of Theorem MLTCV [1089]. Paraphrasing, this theorem says that we can "push" linear transformations "down into" linear combinations, or "pull" linear transformations "up out" of linear combinations. We'll have opportunities to both push and pull.

Theorem LTLC

Linear Transformations and Linear Combinations

Suppose that $T: U \to V$ is a linear transformation, $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t$ are vectors from U and $a_1, a_2, a_3, \ldots, a_t$ are scalars from \mathbb{C} . Then

$$T(a_{1}\mathbf{u}_{1} + a_{2}\mathbf{u}_{2} + a_{3}\mathbf{u}_{3} + \dots + a_{t}\mathbf{u}_{t}) = a_{1}T(\mathbf{u}_{1}) + a_{2}T(\mathbf{u}_{2}) + a_{3}T(\mathbf{u}_{3}) + \dots + a_{t}T(\mathbf{u}_{t})$$

Proof

$$T (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_t \mathbf{u}_t)$$

= $T (a_1 \mathbf{u}_1) + T (a_2 \mathbf{u}_2) + T (a_3 \mathbf{u}_3) + \dots + T (a_t \mathbf{u}_t)$ Definition LT [1072]
= $a_1 T (\mathbf{u}_1) + a_2 T (\mathbf{u}_2) + a_3 T (\mathbf{u}_3) + \dots + a_t T (\mathbf{u}_t)$ Definition LT [1072]

Some authors, especially in more advanced texts, take the conclusion of Theorem LTLC [1093] as the defining condition of a linear transformation. This has the appeal of being a single condition, rather than the two-part condition of Definition LT [1072]. (See Exercise LT.T20 [1118]).

Our next theorem says, informally, that it is enough to know how a linear transformation behaves for inputs from any basis of the domain, and *all* the other outputs are described by a linear combination of these few values. Again, the statement of the theorem, and its proof, are not remarkable, but the insight that goes along with it is very fundamental.

Theorem LTDB

Linear Transformation Defined on a Basis

Suppose $B = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is a basis for the vector space U and $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n$ is a list of vectors from the vector space V (which are not necessarily distinct). Then there is a unique linear transformation, $T: U \to V$, such that $T(\mathbf{u}_i) = \mathbf{v}_i, 1 \le i \le n$.

Proof To prove the existence of T, we construct a function and show that it is a linear transformation (Technique C [1618]). Suppose $\mathbf{w} \in U$ is an arbitrary element of the domain. Then by Theorem VRRB [755] there are unique scalars $a_1, a_2, a_3, \ldots, a_n$ such that

$$\mathbf{w} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_n\mathbf{u}_n$$

Then *define*

$$T(\mathbf{w}) = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \dots + a_n\mathbf{v}_n$$

It should be clear that T behaves as required for n inputs from B. Since the scalars provided by Theorem VRRB [755] are unique, there is no ambiguity in this definition, and T qualifies as a function with domain U and codomain V (i.e. T is well-defined). But is T a linear transformation as well? Let $\mathbf{x} \in U$ be a second element of the domain, and suppose the scalars provided by Theorem VRRB [755] (relative to B) are $b_1, b_2, b_3, \ldots, b_n$. Then

$$T (\mathbf{w} + \mathbf{x}) = T (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_n \mathbf{u}_n + b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2 + \dots + b_n \mathbf{u}_n)$$

= $T ((a_1 + b_1) \mathbf{u}_1 + (a_2 + b_2) \mathbf{u}_2 + \dots + (a_n + b_n) \mathbf{u}_n)$ Definition VS [659]
= $(a_1 + b_1) \mathbf{v}_1 + (a_2 + b_2) \mathbf{v}_2 + \dots + (a_n + b_n) \mathbf{v}_n$ Definition of T
= $a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_n \mathbf{v}_n + b_1 \mathbf{v}_1 + b_2 \mathbf{v}_2 + \dots + b_n \mathbf{v}_n$ Definition VS [659]
= $T (\mathbf{w}) + T (\mathbf{x})$

Let $\alpha \in \mathbb{C}$ be any scalar. Then

$$T (\alpha \mathbf{w}) = T (\alpha (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_n \mathbf{u}_n))$$

= $T (\alpha a_1 \mathbf{u}_1 + \alpha a_2 \mathbf{u}_2 + \alpha a_3 \mathbf{u}_3 + \dots + \alpha a_n \mathbf{u}_n)$ Definition VS [659]
= $\alpha a_1 \mathbf{v}_1 + \alpha a_2 \mathbf{v}_2 + \alpha a_3 \mathbf{v}_3 + \dots + \alpha a_n \mathbf{v}_n$ Definition of T
= $\alpha (a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_n \mathbf{v}_n)$ Definition VS [659]
= $\alpha T (\mathbf{w})$

So by Definition LT [1072], T is a linear transformation.

Is T unique (among all linear transformations that take the \mathbf{u}_i to the \mathbf{v}_i)? Applying Technique U [1625], we posit the existence of a second linear transformation, $S: U \to V$ such that $S(\mathbf{u}_i) = \mathbf{v}_i$, $1 \le i \le n$. Again, let $\mathbf{w} \in U$ represent an arbitrary element of U and let $a_1, a_2, a_3, \ldots, a_n$ be the scalars provided by Theorem VRRB [755] (relative to B). We have,

$$T (\mathbf{w}) = T (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_n \mathbf{u}_n)$$
Theorem VRRB [755]
$$= a_1 T (\mathbf{u}_1) + a_2 T (\mathbf{u}_2) + a_3 T (\mathbf{u}_3) + \dots + a_n T (\mathbf{u}_n)$$
Theorem LTLC [1093]
$$= a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 + \dots + a_n \mathbf{v}_n$$
Definition of T
$$= a_1 S (\mathbf{u}_1) + a_2 S (\mathbf{u}_2) + a_3 S (\mathbf{u}_3) + \dots + a_n S (\mathbf{u}_n)$$
Definition of S
$$= S (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_n \mathbf{u}_n)$$
Theorem LTLC [1093]
$$= S (\mathbf{w})$$
Theorem VRRB [755]

So the output of T and S agree on every input, which means they are equal as functions, T = S. So T is unique.

You might recall facts from analytic geometry, such as "any two points determine a line" and "any three non-collinear points determine a parabola." Theorem LTDB [1094] has much of the same feel. By specifying the n outputs for inputs from a basis, an entire linear transformation is determined. The analogy is not perfect, but the style of these facts are not very dissimilar from Theorem LTDB [1094].

Notice that the statement of Theorem LTDB [1094] asserts the *existence* of a linear transformation with certain properties, while the proof shows us exactly how to define the desired linear transformation. The next examples how to work with linear transformations that we find this way.

Example LTDB1

Linear transformation defined on a basis

Consider the linear transformation $T: \mathbb{C}^3 \to \mathbb{C}^2$ that is required to have the following three values,

$$T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}2\\1\end{bmatrix} \qquad T\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}-1\\4\end{bmatrix} \qquad T\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right) = \begin{bmatrix}6\\0\end{bmatrix}$$

Because

$$B = \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\}$$

is a basis for \mathbb{C}^3 (Theorem SUVB [779]), Theorem LTDB [1094] says there is a unique linear transformation T that behaves this way. How do we compute other values of T? Consider the input

$$\mathbf{w} = \begin{bmatrix} 2\\ -3\\ 1 \end{bmatrix} = (2) \begin{bmatrix} 1\\ 0\\ 0 \end{bmatrix} + (-3) \begin{bmatrix} 0\\ 1\\ 0 \end{bmatrix} + (1) \begin{bmatrix} 0\\ 0\\ 1 \end{bmatrix}$$

Then

$$T(\mathbf{w}) = (2) \begin{bmatrix} 2\\1 \end{bmatrix} + (-3) \begin{bmatrix} -1\\4 \end{bmatrix} + (1) \begin{bmatrix} 6\\0 \end{bmatrix} = \begin{bmatrix} 13\\-10 \end{bmatrix}$$

Doing it again,

$$\mathbf{x} = \begin{bmatrix} 5\\2\\-3 \end{bmatrix} = (5) \begin{bmatrix} 1\\0\\0 \end{bmatrix} + (2) \begin{bmatrix} 0\\1\\0 \end{bmatrix} + (-3) \begin{bmatrix} 0\\0\\1 \end{bmatrix}$$
$$T(\mathbf{x}) = (5) \begin{bmatrix} 2\\1 \end{bmatrix} + (2) \begin{bmatrix} -1\\4 \end{bmatrix} + (-3) \begin{bmatrix} 6\\0 \end{bmatrix} = \begin{bmatrix} -10\\13 \end{bmatrix}$$

 \mathbf{SO}

Any other value of T could be computed in a similar manner. So rather than being given a *formula* for the outputs of T, the *requirement* that T behave in a certain way for the inputs chosen from a basis of the domain, is as sufficient as a formula for computing any value of the function. You might notice some parallels between this example and Example MOLT [1091] or Theorem MLTCV [1089].

Example LTDB2

Linear transformation defined on a basis

Consider the linear transformation $R: \mathbb{C}^3 \to \mathbb{C}^2$ with the three values,

$$R\left(\begin{bmatrix}1\\2\\1\end{bmatrix}\right) = \begin{bmatrix}5\\-1\end{bmatrix} \qquad R\left(\begin{bmatrix}-1\\5\\1\end{bmatrix}\right) = \begin{bmatrix}0\\4\end{bmatrix} \qquad R\left(\begin{bmatrix}3\\1\\4\end{bmatrix}\right) = \begin{bmatrix}2\\3\end{bmatrix}$$

You can check that

$$D = \left\{ \begin{bmatrix} 1\\2\\1 \end{bmatrix}, \begin{bmatrix} -1\\5\\1 \end{bmatrix}, \begin{bmatrix} 3\\1\\4 \end{bmatrix} \right\}$$

is a basis for \mathbb{C}^3 (make the vectors the columns of a square matrix and check that the matrix is nonsingular, Theorem CNMB [789]). By Theorem LTDB [1094] we know there is a unique linear transformation R with the three specified outputs. However, we have to work just a bit harder to take an input vector and express it as a linear combination of the vectors in D. For example, consider,

$$\mathbf{y} = \begin{bmatrix} 8\\ -3\\ 5 \end{bmatrix}$$

Then we must first write \mathbf{y} as a linear combination of the vectors in D and solve for the unknown scalars, to arrive at

$$\mathbf{y} = \begin{bmatrix} 8\\ -3\\ 5 \end{bmatrix} = (3) \begin{bmatrix} 1\\ 2\\ 1 \end{bmatrix} + (-2) \begin{bmatrix} -1\\ 5\\ 1 \end{bmatrix} + (1) \begin{bmatrix} 3\\ 1\\ 4 \end{bmatrix}$$

Then the proof of Theorem LTDB [1094] gives us

$$R\left(\mathbf{y}\right) = (3) \begin{bmatrix} 5\\-1 \end{bmatrix} + (-2) \begin{bmatrix} 0\\4 \end{bmatrix} + (1) \begin{bmatrix} 2\\3 \end{bmatrix} = \begin{bmatrix} 17\\-8 \end{bmatrix}$$

Any other value of R could be computed in a similar manner.

Here is a third example of a linear transformation defined by its action on a basis, only with more abstract vector spaces involved.

 \boxtimes

Example LTDB3

Linear transformation defined on a basis

The set $W = \{p(x) \in P_3 \mid p(1) = 0, p(3) = 0\} \subseteq P_3$ is a subspace of the vector space of polynomials P_3 . This subspace has $C = \{3 - 4x + x^2, 12 - 13x + x^3\}$ as a basis (check this!). Suppose we consider the linear transformation $S: P_3 \to M_{22}$ with values

$$S(3 - 4x + x^{2}) = \begin{bmatrix} 1 & -3\\ 2 & 0 \end{bmatrix} \qquad S(12 - 13x + x^{3}) = \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix}$$

By Theorem LTDB [1094] we know there is a unique linear transformation with these two values. To illustrate a sample computation of S, consider $q(x) = 9 - 6x - 5x^2 + 2x^3$. Verify that q(x) is an element of W (does it have roots at x = 1 and x = 3?), then find the scalars needed to write it as a linear combination of the basis vectors in C. Because

$$q(x) = 9 - 6x - 5x^{2} + 2x^{3} = (-5)(3 - 4x + x^{2}) + (2)(12 - 13x + x^{3})$$

The proof of Theorem LTDB [1094] gives us

$$S(q) = (-5) \begin{bmatrix} 1 & -3 \\ 2 & 0 \end{bmatrix} + (2) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -5 & 17 \\ -8 & 0 \end{bmatrix}$$

And all the other outputs of S could be computed in the same manner. Every output of S will have a zero in the second row, second column. Can you see why this is so?

Informally, we can describe Theorem LTDB [1094] by saying "it is enough to know what a linear transformation does to a basis (of the domain)."

Subsection PI Pre-Images

The definition of a function requires that for each input in the domain there is *exactly* one output in the codomain. However, the correspondence does not have to behave the other way around. A member of the codomain might have many inputs from the domain that create it, or it may have none at all. To formalize our discussion of this aspect of linear transformations, we define the pre-image.

Definition PI Pre-Image

Suppose that $T: U \to V$ is a linear transformation. For each **v**, define the **pre-image** of **v** to be the subset of U given by

$$T^{-1}(\mathbf{v}) = \{ \mathbf{u} \in U | T(\mathbf{u}) = \mathbf{v} \}$$

 \triangle

In other words, $T^{-1}(\mathbf{v})$ is the set of all those vectors in the domain U that get "sent" to the vector \mathbf{v} .

Example SPIAS Sample pre-images, Archetype S

Archetype S [1785] is the linear transformation defined by

$$T: \mathbb{C}^3 \to M_{22}, \quad T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) = \begin{bmatrix} a-b & 2a+2b+c \\ 3a+b+c & -2a-6b-2c \end{bmatrix}$$

We could compute a pre-image for every element of the codomain M_{22} . However, even in a free textbook, we do not have the room to do that, so we will compute just two.

Choose

$$\mathbf{v} = \begin{bmatrix} 2 & 1\\ 3 & 2 \end{bmatrix} \in M_{22}$$

for no particular reason. What is $T^{-1}(\mathbf{v})$? Suppose $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \in T^{-1}(\mathbf{v})$. The condition that $T(\mathbf{u}) = \mathbf{v}$ becomes

$$\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} = \mathbf{v} = T(\mathbf{u}) = T\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \right) = \begin{bmatrix} u_1 - u_2 & 2u_1 + 2u_2 + u_3 \\ 3u_1 + u_2 + u_3 & -2u_1 - 6u_2 - 2u_3 \end{bmatrix}$$

Using matrix equality (Definition ME [427]), we arrive at a system of four equations in the three unknowns u_1 , u_2 , u_3 with an augmented matrix that we can row-reduce in the hunt

for solutions,

$$\begin{bmatrix} 1 & -1 & 0 & 2\\ 2 & 2 & 1 & 1\\ 3 & 1 & 1 & 3\\ -2 & -6 & -2 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & \frac{1}{4} & \frac{5}{4}\\ 0 & 1 & \frac{1}{4} & -\frac{3}{4}\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We recognize this system as having infinitely many solutions described by the single free variable u_3 . Eventually obtaining the vector form of the solutions (Theorem VFSLS [244]), we can describe the preimage precisely as,

$$T^{-1} (\mathbf{v}) = \left\{ \mathbf{u} \in \mathbb{C}^3 \middle| T(\mathbf{u}) = \mathbf{v} \right\}$$
$$= \left\{ \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \middle| u_1 = \frac{5}{4} - \frac{1}{4}u_3, u_2 = -\frac{3}{4} - \frac{1}{4}u_3 \right\}$$
$$= \left\{ \begin{bmatrix} \frac{5}{4} - \frac{1}{4}u_3 \\ -\frac{3}{4} - \frac{1}{4}u_3 \\ u_3 \end{bmatrix} \middle| u_3 \in \mathbb{C}^3 \right\}$$
$$= \left\{ \begin{bmatrix} \frac{5}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + u_3 \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \\ 1 \end{bmatrix} \middle| u_3 \in \mathbb{C}^3 \right\}$$
$$= \begin{bmatrix} \frac{5}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + \left\langle \left\{ \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \\ 1 \end{bmatrix} \right\} \right\rangle$$

This last line is merely a suggestive way of describing the set on the previous line. You might create three or four vectors in the preimage, and evaluate T with each. Was the result what you expected? For a hint of things to come, you might try evaluating T with just the lone vector in the spanning set above. What was the result? Now take a look back at Theorem PSPHS [256]. Hmmmm.

OK, let's compute another preimage, but with a different outcome this time. Choose

$$\mathbf{v} = \begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} \in M_{22}$$

What is $T^{-1}(\mathbf{v})$? Suppose $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \in T^{-1}(\mathbf{v})$. That $T(\mathbf{u}) = \mathbf{v}$ becomes
$$\begin{bmatrix} 1 & 1 \\ 2 & 4 \end{bmatrix} = \mathbf{v} = T(\mathbf{u}) = T\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \right) = \begin{bmatrix} u_1 - u_2 & 2u_1 + 2u_2 + u_3 \\ 3u_1 + u_2 + u_3 & -2u_1 - 6u_2 - 2u_3 \end{bmatrix}$$

Using matrix equality (Definition ME [427]), we arrive at a system of four equations in the three unknowns u_1 , u_2 , u_3 with an augmented matrix that we can row-reduce in the hunt for solutions,

$$\begin{bmatrix} 1 & -1 & 0 & 1 \\ 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 2 \\ -2 & -6 & -2 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & \frac{1}{4} & 0 \\ 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem RCLS [120] we recognize this system as inconsistent. So no vector **u** is a member of $T^{-1}(\mathbf{v})$ and so

$$T^{-1}\left(\mathbf{v}\right) = \emptyset$$

 \boxtimes

The preimage is just a set, it is almost never a subspace of U (you might think about just when $T^{-1}(\mathbf{v})$ is a subspace, see Exercise ILT.T10 [1158]). We will describe its properties going forward, and it will be central to the main ideas of this chapter.

Subsection NLTFO New Linear Transformations From Old

We can combine linear transformations in natural ways to create new linear transformations. So we will define these combinations and then prove that the results really are still linear transformations. First the sum of two linear transformations.

Definition LTA

Linear Transformation Addition

Suppose that $T: U \to V$ and $S: U \to V$ are two linear transformations with the same domain and codomain. Then their **sum** is the function $T + S: U \to V$ whose outputs are defined by

$$(T+S)\left(\mathbf{u}\right) = T\left(\mathbf{u}\right) + S\left(\mathbf{u}\right)$$

 \triangle

Notice that the first plus sign in the definition is the operation being defined, while the second one is the vector addition in V. (Vector addition in U will appear just now in the proof that T + S is a linear transformation.) Definition LTA [1105] only provides a function. It would be nice to know that when the constituents (T, S) are linear transformations, then so too is T + S.

Theorem SLTLT

Sum of Linear Transformations is a Linear Transformation

Suppose that $T: U \to V$ and $S: U \to V$ are two linear transformations with the same domain and codomain. Then $T + S: U \to V$ is a linear transformation. \Box

Proof We simply check the defining properties of a linear transformation (Definition LT [1072]). This is a good place to consistently ask yourself which objects are being combined with which operations.

$$(T + S) (\mathbf{x} + \mathbf{y}) = T (\mathbf{x} + \mathbf{y}) + S (\mathbf{x} + \mathbf{y})$$

= T (\mathbf{x}) + T (\mathbf{y}) + S (\mathbf{x}) + S (\mathbf{y})
= T (\mathbf{x}) + S (\mathbf{x}) + T (\mathbf{y}) + S (\mathbf{y})
= (T + S) (\mathbf{x}) + (T + S) (\mathbf{y})

Definition LTA [1105] Definition LT [1072] Property C [660] in V Definition LTA [1105]

and

$$(T+S)(\alpha \mathbf{x}) = T(\alpha \mathbf{x}) + S(\alpha \mathbf{x})$$

Definition LTA [1105]

$$= \alpha T (\mathbf{x}) + \alpha S (\mathbf{x})$$

Definition LT [1072]
$$= \alpha (T (\mathbf{x}) + S (\mathbf{x}))$$

Property DVA [660] in V
$$= \alpha (T + S) (\mathbf{x})$$

Definition LTA [1105]

Example STLT Sum of two linear transformations

Suppose that $T: \mathbb{C}^2 \to \mathbb{C}^3$ and $S: \mathbb{C}^2 \to \mathbb{C}^3$ are defined by

$$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}x_1+2x_2\\3x_1-4x_2\\5x_1+2x_2\end{bmatrix} \qquad \qquad S\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}4x_1-x_2\\x_1+3x_2\\-7x_1+5x_2\end{bmatrix}$$

Then by Definition LTA [1105], we have

$$(T+S)\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) + S\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}x_1+2x_2\\3x_1-4x_2\\5x_1+2x_2\end{bmatrix} + \begin{bmatrix}4x_1-x_2\\x_1+3x_2\\-7x_1+5x_2\end{bmatrix} = \begin{bmatrix}5x_1+x_2\\4x_1-x_2\\-2x_1+7x_2\end{bmatrix}$$

and by Theorem SLTLT [1106] we know T + S is also a linear transformation from \mathbb{C}^2 to \mathbb{C}^3 .

Definition LTSM

Linear Transformation Scalar Multiplication

Suppose that $T: U \to V$ is a linear transformation and $\alpha \in \mathbb{C}$. Then the scalar multiple is the function $\alpha T: U \to V$ whose outputs are defined by

$$\left(\alpha T\right)\left(\mathbf{u}\right) = \alpha T\left(\mathbf{u}\right)$$

 \triangle

Given that T is a linear transformation, it would be nice to know that αT is also a linear transformation.

Theorem MLTLT

Multiple of a Linear Transformation is a Linear Transformation

Suppose that $T: U \to V$ is a linear transformation and $\alpha \in \mathbb{C}$. Then $(\alpha T): U \to V$ is a linear transformation.

Proof We simply check the defining properties of a linear transformation (Definition LT [1072]). This is another good place to consistently ask yourself which objects are being combined with which operations.

$$(\alpha T) (\mathbf{x} + \mathbf{y}) = \alpha (T (\mathbf{x} + \mathbf{y}))$$
Definition LTSM [1108]
$$= \alpha (T (\mathbf{x}) + T (\mathbf{y}))$$
Definition LT [1072]
$$= \alpha T (\mathbf{x}) + \alpha T (\mathbf{y})$$
Property DVA [660] in V
$$= (\alpha T) (\mathbf{x}) + (\alpha T) (\mathbf{y})$$
Definition LTSM [1108]

and

$$(\alpha T) (\beta \mathbf{x}) = \alpha T (\beta \mathbf{x})$$
$$= \alpha (\beta T (\mathbf{x}))$$
$$= (\alpha \beta) T (\mathbf{x})$$
$$= (\beta \alpha) T (\mathbf{x})$$
$$= \beta (\alpha T (\mathbf{x}))$$
$$= \beta ((\alpha T) (\mathbf{x}))$$

Definition LTSM [1108] Definition LT [1072] Property SMA [660] in VCommutativity in \mathbb{C} Property SMA [660] in VDefinition LTSM [1108]

Example SMLT Scalar multiple of a linear transformation Suppose that $T: \mathbb{C}^4 \to \mathbb{C}^3$ is defined by

$$T\left(\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}\right) = \begin{bmatrix}x_1 + 2x_2 - x_3 + 2x_4\\x_1 + 5x_2 - 3x_3 + x_4\\-2x_1 + 3x_2 - 4x_3 + 2x_4\end{bmatrix}$$

For the sake of an example, choose $\alpha = 2$, so by Definition LTSM [1108], we have

$$\alpha T \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = 2T \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = 2 \begin{bmatrix} x_1 + 2x_2 - x_3 + 2x_4 \\ x_1 + 5x_2 - 3x_3 + x_4 \\ -2x_1 + 3x_2 - 4x_3 + 2x_4 \end{bmatrix} = \begin{bmatrix} 2x_1 + 4x_2 - 2x_3 + 4x_4 \\ 2x_1 + 10x_2 - 6x_3 + 2x_4 \\ -4x_1 + 6x_2 - 8x_3 + 4x_4 \end{bmatrix}$$

and by Theorem MLTLT [1108] we know 2T is also a linear transformation from \mathbb{C}^4 to \mathbb{C}^3 .

Now, let's imagine we have two vector spaces, U and V, and we collect every possible linear transformation from U to V into one big set, and call it $\mathcal{L}T(U, V)$. Definition LTA [1105] and Definition LTSM [1108] tell us how we can "add" and "scalar multiply" two elements of $\mathcal{L}T(U, V)$. Theorem SLTLT [1106] and Theorem MLTLT [1108] tell us that if we do these operations, then the resulting functions are linear transformations that are also in $\mathcal{L}T(U, V)$. Hmmmm, sounds like a vector space to me! A set of objects, an addition and a scalar multiplication. Why not?

Theorem VSLT

Vector Space of Linear Transformations

Suppose that U and V are vector spaces. Then the set of all linear transformations from U to V, $\mathcal{L}T(U, V)$ is a vector space when the operations are those given in Definition LTA [1105] and Definition LTSM [1108].

Proof Theorem SLTLT [1106] and Theorem MLTLT [1108] provide two of the ten properties in Definition VS [659]. However, we still need to verify the remaining eight properties. By

and large, the proofs are straightforward and rely on concocting the obvious object, or by reducing the question to the same vector space property in the vector space V.

The zero vector is of some interest, though. What linear transformation would we add to any other linear transformation, so as to keep the second one unchanged? The answer is $Z: U \to V$ defined by $Z(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Notice how we do not need to know any of the specifics about U and V to make this definition of Z.

Definition LTC Linear Transformation Composition

Suppose that $T: U \to V$ and $S: V \to W$ are linear transformations. Then the **composition** of S and T is the function $(S \circ T): U \to W$ whose outputs are defined by

$$(S \circ T) (\mathbf{u}) = S (T (\mathbf{u}))$$

Λ	
L	7

Given that T and S are linear transformations, it would be nice to know that $S \circ T$ is also a linear transformation.

Theorem CLTLT

Composition of Linear Transformations is a Linear Transformation

Suppose that $T: U \to V$ and $S: V \to W$ are linear transformations. Then $(S \circ T): U \to W$ is a linear transformation.

Proof We simply check the defining properties of a linear transformation (Definition LT

[1072]).

$$(S \circ T) (\mathbf{x} + \mathbf{y}) = S (T (\mathbf{x} + \mathbf{y}))$$

= $S (T (\mathbf{x}) + T (\mathbf{y}))$
= $S (T (\mathbf{x})) + S (T (\mathbf{y}))$
= $(S \circ T) (\mathbf{x}) + (S \circ T) (\mathbf{y})$

Definition LTC [1111] Definition LT [1072] for TDefinition LT [1072] for SDefinition LTC [1111]

and

$$(S \circ T) (\alpha \mathbf{x}) = S (T (\alpha \mathbf{x}))$$
$$= S (\alpha T (\mathbf{x}))$$
$$= \alpha S (T (\mathbf{x}))$$
$$= \alpha (S \circ T) (\mathbf{x})$$

Definition LTC [1111] Definition LT [1072] for TDefinition LT [1072] for SDefinition LTC [1111]

Example CTLT Composition of two linear transformations Suppose that $T: \mathbb{C}^2 \to \mathbb{C}^4$ and $S: \mathbb{C}^4 \to \mathbb{C}^3$ are defined by

$$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}x_1 + 2x_2\\3x_1 - 4x_2\\5x_1 + 2x_2\\6x_1 - 3x_2\end{bmatrix} \qquad S\left(\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}\right) = \begin{bmatrix}2x_1 - x_2 + x_3 - x_4\\5x_1 - 3x_2 + 8x_3 - 2x_4\\-4x_1 + 3x_2 - 4x_3 + 5x_4\end{bmatrix}$$

Then by Definition LTC [1111]

$$(S \circ T) \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = S \left(T \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) \right)$$
$$= S \left(\begin{bmatrix} x_1 + 2x_2 \\ 3x_1 - 4x_2 \\ 5x_1 + 2x_2 \\ 6x_1 - 3x_2 \end{bmatrix} \right)$$
$$= \begin{bmatrix} 2(x_1 + 2x_2) - (3x_1 - 4x_2) + (5x_1 + 2x_2) - (6x_1 - 3x_2) \\ 5(x_1 + 2x_2) - 3(3x_1 - 4x_2) + 8(5x_1 + 2x_2) - 2(6x_1 - 3x_2) \\ -4(x_1 + 2x_2) + 3(3x_1 - 4x_2) - 4(5x_1 + 2x_2) + 5(6x_1 - 3x_2) \end{bmatrix}$$
$$= \begin{bmatrix} -2x_1 + 13x_2 \\ 24x_1 + 44x_2 \\ 15x_1 - 43x_2 \end{bmatrix}$$

and by Theorem CLTLT [1111] $S \circ T$ is a linear transformation from \mathbb{C}^2 to \mathbb{C}^3 .

 \boxtimes

Here is an interesting exercise that will presage an important result later. In Example STLT [1107] compute (via Theorem MLTCV [1089]) the matrix of T, S and T + S. Do you see a relationship between these three matrices?

In Example SMLT [1109] compute (via Theorem MLTCV [1089]) the matrix of T and 2T. Do you see a relationship between these two matrices?

Here's the tough one. In Example CTLT [1112] compute (via Theorem MLTCV [1089]) the matrix of T, S and $S \circ T$. Do you see a relationship between these three matrices???

Subsection READ Reading Questions

1. Is the function below a linear transformation? Why or why not?

$$T: \mathbb{C}^3 \to \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} 3x_1 - x_2 + x_3 \\ 8x_2 - 6 \end{bmatrix}$$

2. Determine the matrix representation of the linear transformation S below.

$$S: \mathbb{C}^2 \to \mathbb{C}^3, \quad S\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}3x_1 + 5x_2\\8x_1 - 3x_2\\-4x_1\end{bmatrix}$$

3. Theorem LTLC [1093] has a fairly simple proof. Yet the result itself is very powerful. Comment on why we might say this.

Subsection EXC Exercises

C15 The archetypes below are all linear transformations whose domains and codomains are vector spaces of column vectors (Definition VSCV [198]). For each one, compute the matrix representation described in the proof of Theorem MLTCV [1089].

Archetype M [1750] Archetype N [1755] Archetype O [1760] Archetype P [1766] Archetype Q [1771] Archetype R [1779] Contributed by Robert Beezer

C16 Find the matrix representation of
$$T: \mathbb{C}^3 \to \mathbb{C}^4$$
 given by $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 3x + 2y + z \\ x + y + z \\ x - 3y \\ 2x + 3y + z \end{bmatrix}$

Contributed by Chris Black Solution [1120]

C20 Let $\mathbf{w} = \begin{bmatrix} -3\\ 1\\ 4 \end{bmatrix}$. Referring to Example MOLT [1091], compute $S(\mathbf{w})$ two different ways. First use the definition of S, then compute the matrix-vector product $C\mathbf{w}$ (Definition MVP [461]).

Contributed by Robert Beezer Solution [1120]

C25 Define the linear transformation

$$T: \mathbb{C}^3 \to \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1\\x_2\\x_3 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 - x_2 + 5x_3\\-4x_1 + 2x_2 - 10x_3 \end{bmatrix}$$

Verify that T is a linear transformation. Contributed by Robert Beezer Solution [1120]

C26 Verify that the function below is a linear transformation.

$$T: P_2 \to \mathbb{C}^2, \quad T\left(a + bx + cx^2\right) = \begin{bmatrix} 2a - b \\ b + c \end{bmatrix}$$

Contributed by Robert Beezer Solution [1121]

C30 Define the linear transformation

$$T: \mathbb{C}^3 \to \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1\\x_2\\x_3 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 - x_2 + 5x_3\\-4x_1 + 2x_2 - 10x_3 \end{bmatrix}$$

Compute the preimages, $T^{-1}\left(\begin{bmatrix} 2\\ 3 \end{bmatrix} \right)$ and $T^{-1}\left(\begin{bmatrix} 4\\ -8 \end{bmatrix} \right)$. Contributed by Robert Beezer Solution [1122]

C31 For the linear transformation S compute the pre-images.

$$S: \mathbb{C}^3 \to \mathbb{C}^3, \quad S\left(\begin{bmatrix}a\\b\\c\end{bmatrix}\right) = \begin{bmatrix}a-2b-c\\3a-b+2c\\a+b+2c\end{bmatrix}$$
$$S^{-1}\left(\begin{bmatrix}-2\\5\\3\end{bmatrix}\right) \qquad \qquad S^{-1}\left(\begin{bmatrix}-5\\5\\7\end{bmatrix}\right)$$

Contributed by Robert Beezer Solution [1123] **C40** If $T: \mathbb{C}^2 \to \mathbb{C}^2$ satisfies $T\left(\begin{bmatrix} 2\\1 \end{bmatrix} \right) = \begin{bmatrix} 3\\4 \end{bmatrix}$ and $T\left(\begin{bmatrix} 1\\1 \end{bmatrix} \right) = \begin{bmatrix} -1\\2 \end{bmatrix}$, find $T\left(\begin{bmatrix} 4\\3 \end{bmatrix} \right)$. Contributed by Chris Black Solution [1124]

C41 If
$$T: \mathbb{C}^2 \to \mathbb{C}^3$$
 satisfies $T\left(\begin{bmatrix}2\\3\end{bmatrix}\right) = \begin{bmatrix}2\\2\\1\end{bmatrix}$ and $T\left(\begin{bmatrix}3\\4\end{bmatrix}\right) = \begin{bmatrix}-1\\0\\2\end{bmatrix}$, find the matrix

representation of T.

Contributed by Chris Black Solution [1125]

C42 Define $T: M_{2,2} \to \mathbb{R}$ by $T\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right) = a + b + c - d$. Find the pre-image $T^{-1}(3)$. Contributed by Chris Black Solution [1126]

C43 Define $T: P_3 \to P_2$ by $T(a + bx + cx^2 + dx^3) = b + 2cx + 3dx^2$. Find the pre-image of **0**. Does this linear transformation seem familiar? Contributed by Chris Black Solution [1126]

M10 Define two linear transformations, $T \colon \mathbb{C}^4 \to \mathbb{C}^3$ and $S \colon \mathbb{C}^3 \to \mathbb{C}^2$ by

$$S\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}x_1 - 2x_2 + 3x_3\\5x_1 + 4x_2 + 2x_3\end{bmatrix} \qquad T\left(\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}\right) = \begin{bmatrix}-x_1 + 3x_2 + x_3 + 9x_4\\2x_1 + x_3 + 7x_4\\4x_1 + 2x_2 + x_3 + 2x_4\end{bmatrix}$$

Using the proof of Theorem MLTCV [1089] compute the matrix representations of the three linear transformations T, S and $S \circ T$. Discover and comment on the relationship between these three matrices.

Contributed by Robert Beezer Solution [1127]

M60 Suppose U and V are vector spaces and define a function $Z: U \to V$ by $T(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Prove that Z is a (stupid) linear transformation. (See Exercise ILT.M60 [1158], Exercise SLT.M60 [1201], Exercise IVLT.M60 [1248].) Contributed by Robert Beezer

T20 Use the conclusion of Theorem LTLC [1093] to motivate a new definition of a linear

transformation. Then prove that your new definition is equivalent to Definition LT [1072]. (Technique D [1611] and Technique E [1619] might be helpful if you are not sure what you are being asked to prove here.) Contributed by Robert Beezer

Subsection SOL Solutions

C16 Contributed by Chris Black Statement [1115]
Answer:
$$A_T = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & -3 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$
.
C20 Contributed by Robert Beezer Statement [1116]
 $\begin{bmatrix} 9 \\ 9 \end{bmatrix}$

In both cases the result will be $S(\mathbf{w}) = \begin{vmatrix} 2 \\ -9 \\ 4 \end{vmatrix}$.

C25Contributed by Robert BeezerStatement [1116]We can rewrite T as follows:

$$T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}2x_1 - x_2 + 5x_3\\-4x_1 + 2x_2 - 10x_3\end{bmatrix} = x_1\begin{bmatrix}2\\-4\end{bmatrix} + x_2\begin{bmatrix}-1\\2\end{bmatrix} + x_3\begin{bmatrix}5\\-10\end{bmatrix} = \begin{bmatrix}2 & -1 & 5\\-4 & 2 & -10\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

and Theorem MBLT [1086] tell us that any function of this form is a linear transformation.

C26 Contributed by Robert Beezer Statement [1116] Check the two conditions of Definition LT [1072].

$$T (\mathbf{u} + \mathbf{v}) = T \left(\left(a + bx + cx^2 \right) + \left(d + ex + fx^2 \right) \right)$$
$$= T \left(\left(a + d \right) + \left(b + e \right) x + \left(c + f \right) x^2 \right)$$
$$= \left[2(a + d) - (b + e) \\ (b + e) + (c + f) \right]$$
$$= \left[(2a - b) + (2d - e) \\ (b + c) + (e + f) \right]$$
$$= \left[2a - b \\ b + c \right] + \left[2d - e \\ e + f \right]$$
$$= T (\mathbf{u}) + T (\mathbf{v})$$

and

$$T (\alpha \mathbf{u}) = T (\alpha (a + bx + cx^{2}))$$

= $T ((\alpha a) + (\alpha b) x + (\alpha c) x^{2})$
= $\begin{bmatrix} 2(\alpha a) - (\alpha b) \\ (\alpha b) + (\alpha c) \end{bmatrix}$
= $\begin{bmatrix} \alpha(2a - b) \\ \alpha(b + c) \end{bmatrix}$

$$= \alpha \begin{bmatrix} 2a - b \\ b + c \end{bmatrix}$$
$$= \alpha T (\mathbf{u})$$

So T is indeed a linear transformation.

C30 Contributed by Robert Beezer Statement [1116] For the first pre-image, we want $\mathbf{x} \in \mathbb{C}^3$ such that $T(\mathbf{x}) = \begin{bmatrix} 2\\ 3 \end{bmatrix}$. This becomes,

$$\begin{bmatrix} 2x_1 - x_2 + 5x_3 \\ -4x_1 + 2x_2 - 10x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Vector equality gives a system of two linear equations in three variables, represented by the augmented matrix

$$\begin{bmatrix} 2 & -1 & 5 & 2 \\ -4 & 2 & -10 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{5}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

so the system is inconsistent and the pre-image is the empty set. For the second pre-image the same procedure leads to an augmented matrix with a different vector of constants

$$\begin{bmatrix} 2 & -1 & 5 & 4 \\ -4 & 2 & -10 & -8 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{5}{2} & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This system is consistent and has infinitely many solutions, as we can see from the presence

of the two free variables $(x_2 \text{ and } x_3)$ both to zero. We apply Theorem VFSLS [244] to obtain

$$T^{-1}\left(\begin{bmatrix}4\\-8\end{bmatrix}\right) = \left\{ \begin{bmatrix}2\\0\\0\end{bmatrix} + x_2\begin{bmatrix}\frac{1}{2}\\1\\0\end{bmatrix} + x_3\begin{bmatrix}-\frac{5}{2}\\0\\1\end{bmatrix} \middle| x_2, x_3 \in \mathbb{C} \right\}$$

C31 Contributed by Robert Beezer Statement [1117] We work from the definition of the pre-image, Definition PI [1101]. Setting

$$S\left(\begin{bmatrix}a\\b\\c\end{bmatrix}\right) = \begin{bmatrix}-2\\5\\3\end{bmatrix}$$

we arrive at a system of three equations in three variables, with an augmented matrix that we row-reduce in a search for solutions,

$$\begin{bmatrix} 1 & -2 & -1 & -2 \\ 3 & -1 & 2 & 5 \\ 1 & 1 & 2 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

With a leading 1 in the last column, this system is inconsistent (Theorem RCLS [120]), and there are no values of a, b and c that will create an element of the pre-image. So the preimage is the empty set.

We work from the definition of the pre-image, Definition PI [1101]. Setting

$$S\left(\begin{bmatrix}a\\b\\c\end{bmatrix}\right) = \begin{bmatrix}-5\\5\\7\end{bmatrix}$$

we arrive at a system of three equations in three variables, with an augmented matrix that we row-reduce in a search for solutions,

$$\begin{bmatrix} 1 & -2 & -1 & -5 \\ 3 & -1 & 2 & 5 \\ 1 & 1 & 2 & 7 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The solution set to this system, which is also the desired pre-image, can be expressed using the vector form of the solutions (Theorem VFSLS [244])

$$S^{-1}\left(\begin{bmatrix} -5\\5\\7 \end{bmatrix} \right) = \left\{ \begin{bmatrix} 3\\4\\0 \end{bmatrix} + c \begin{bmatrix} -1\\-1\\1 \end{bmatrix} \middle| c \in \mathbb{C} \right\} = \begin{bmatrix} 3\\4\\0 \end{bmatrix} + \left\langle \left\{ \begin{bmatrix} -1\\-1\\1 \end{bmatrix} \right\} \right\rangle$$

Does the final expression for this set remind you of Theorem KPI [1143]?

C40 Contributed by Chris Black Statement [1117]
Since
$$\begin{bmatrix} 4\\3 \end{bmatrix} = \begin{bmatrix} 2\\1 \end{bmatrix} + 2 \begin{bmatrix} 1\\1 \end{bmatrix}$$
, we have
 $T\left(\begin{bmatrix} 4\\3 \end{bmatrix} \right) = T\left(\begin{bmatrix} 2\\1 \end{bmatrix} + 2 \begin{bmatrix} 1\\1 \end{bmatrix} \right) = T\left(\begin{bmatrix} 2\\1 \end{bmatrix} \right) + 2T\left(\begin{bmatrix} 1\\1 \end{bmatrix} \right)$

$$= \begin{bmatrix} 3\\4 \end{bmatrix} + 2 \begin{bmatrix} -1\\2 \end{bmatrix} = \begin{bmatrix} 1\\8 \end{bmatrix}$$

C41 Contributed by Chris Black Statement [1117]

First, we need to write the standard basis vectors \mathbf{e}_1 and \mathbf{e}_2 as linear combinations of $\begin{vmatrix} 2\\3 \end{vmatrix}$

and
$$\begin{bmatrix} 3\\4 \end{bmatrix}$$
. Starting with \mathbf{e}_1 , we see that $\mathbf{e}_1 = -4\begin{bmatrix} 2\\3 \end{bmatrix} + 3\begin{bmatrix} 3\\4 \end{bmatrix}$, so we have
$$T\left(\mathbf{e}_1\right) = T\left(-4\begin{bmatrix} 2\\3 \end{bmatrix} + 3\begin{bmatrix} 3\\4 \end{bmatrix}\right) = -4T\left(\begin{bmatrix} 2\\3 \end{bmatrix}\right) + 3T\left(\begin{bmatrix} 3\\4 \end{bmatrix}\right)$$

$$= -4 \begin{bmatrix} 2\\2\\1 \end{bmatrix} + 3 \begin{bmatrix} -1\\0\\2 \end{bmatrix} = \begin{bmatrix} -11\\-8\\2 \end{bmatrix}.$$

Repeating the process for \mathbf{e}_2 , we have $\mathbf{e}_2 = 3 \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, and we then see that

$$T(\mathbf{e}_2) = T\left(3\begin{bmatrix}2\\3\end{bmatrix} - 2\begin{bmatrix}3\\4\end{bmatrix}\right) = 3T\left(\begin{bmatrix}2\\3\end{bmatrix}\right) - 2T\left(\begin{bmatrix}3\\4\end{bmatrix}\right)$$
$$= 3\begin{bmatrix}2\\2\\1\end{bmatrix} - 2\begin{bmatrix}-1\\0\\2\end{bmatrix} = \begin{bmatrix}8\\6\\-1\end{bmatrix}.$$

Thus, the matrix representation of T is $A_T = \begin{bmatrix} -11 & 8 \\ -8 & 6 \\ 2 & -1 \end{bmatrix}$.

C42 Contributed by Chris Black Statement [1118] The preimage $T^{-1}(3)$ is the set of all matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ so that $T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = 3$. A matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is in the preimage if a + b + c - d = 3, i.e. d = a + b + c - 3. This is the set. (But the set is *not* a vector space. Why not?)

$$T^{-1}(3) = \left\{ \begin{bmatrix} a & b \\ c & a+b+c-3 \end{bmatrix} \middle| a, b, c \in \mathbb{C} \right\}$$

C43 Contributed by Chris Black Statement [1118]

The preimage $T^{-1}(0)$ is the set of all polynomials $a+bx+cx^2+dx^3$ so that $T(a+bx+cx^2+dx^3) = 0$. Thus, $b+2cx+3dx^2 = 0$, where the 0 represents the zero polynomial. In order to satisfy this equation, we must have b = 0, c = 0, and d = 0. Thus, $T^{-1}(0)$ is precisely the set of all constant polynomials – polynomials of degree 0. Symbolically, this is $T^{-1}(0) = \{a | a \in \mathbb{C}\}$. Does this seem familiar? What other operation sends constant functions to 0?

M10 Contributed by Robert Beezer Statement [1118]

$$\begin{bmatrix} 1 & -2 & 3 \\ 5 & 4 & 2 \end{bmatrix} \begin{bmatrix} -1 & 3 & 1 & 9 \\ 2 & 0 & 1 & 7 \\ 4 & 2 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 7 & 9 & 2 & 1 \\ 11 & 19 & 11 & 77 \end{bmatrix}$$

Section ILT Injective Linear Transformations

Some linear transformations possess one, or both, of two key properties, which go by the names injective and surjective. We will see that they are closely related to ideas like linear independence and spanning, and subspaces like the null space and the column space. In this section we will define an injective linear transformation and analyze the resulting consequences. The next section will do the same for the surjective property. In the final section of this chapter we will see what happens when we have the two properties simultaneously.

As usual, we lead with a definition.

Definition ILT

Injective Linear Transformation

Suppose $T: U \to V$ is a linear transformation. Then T is **injective** if whenever $T(\mathbf{x}) = T(\mathbf{y})$, then $\mathbf{x} = \mathbf{y}$.

Given an arbitrary function, it is possible for two different inputs to yield the same output (think about the function $f(x) = x^2$ and the inputs x = 3 and x = -3). For an injective function, this never happens. If we have equal outputs $(T(\mathbf{x}) = T(\mathbf{y}))$ then we must have achieved those equal outputs by employing equal inputs $(\mathbf{x} = \mathbf{y})$. Some authors prefer the term **one-to-one** where we use injective, and we will sometimes refer to an injective linear transformation as an **injection**.

Subsection EILT Examples of Injective Linear Transformations

It is perhaps most instructive to examine a linear transformation that is not injective first.

Example NIAQ Not injective, Archetype Q

Archetype Q [1771] is the linear transformation

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} -2x_1 + 3x_2 + 3x_3 - 6x_4 + 3x_5 \\ -16x_1 + 9x_2 + 12x_3 - 28x_4 + 28x_5 \\ -19x_1 + 7x_2 + 14x_3 - 32x_4 + 37x_5 \\ -21x_1 + 9x_2 + 15x_3 - 35x_4 + 39x_5 \\ -9x_1 + 5x_2 + 7x_3 - 16x_4 + 16x_5 \end{bmatrix}$$

Notice that for

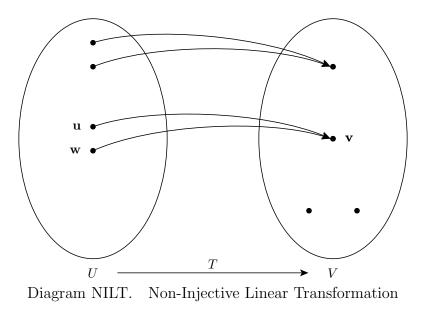
$$\mathbf{x} = \begin{bmatrix} 1\\3\\-1\\2\\4 \end{bmatrix} \qquad \qquad \mathbf{y} = \begin{bmatrix} 4\\7\\0\\5\\7 \end{bmatrix}$$

we have

$$T\left(\begin{bmatrix}1\\3\\-1\\2\\4\end{bmatrix}\right) = \begin{bmatrix}4\\55\\72\\77\\31\end{bmatrix} \qquad T\left(\begin{bmatrix}4\\7\\0\\5\\7\end{bmatrix}\right) = \begin{bmatrix}4\\55\\72\\77\\31\end{bmatrix}$$

So we have two vectors from the domain, $\mathbf{x} \neq \mathbf{y}$, yet $T(\mathbf{x}) = T(\mathbf{y})$, in violation of Definition ILT [1128]. This is another example where you should not concern yourself with how \mathbf{x} and \mathbf{y} were selected, as this will be explained shortly. However, do understand *why* these two vectors provide enough evidence to conclude that T is not injective.

Here's a cartoon of a non-injective linear transformation. Notice that the central feature of this cartoon is that $T(\mathbf{u}) = \mathbf{v} = T(\mathbf{w})$. Even though this happens again with some unnamed vectors, it only takes one occurrence to destroy the possibility of injectivity. Note also that the two vectors displayed in the bottom of V have no bearing, either way, on the injectivity of T.



To show that a linear transformation is not injective, it is enough to find a single pair of inputs that get sent to the identical output, as in Example NIAQ [1129]. However, to show that a linear transformation is injective we must establish that this coincidence of outputs *never* occurs. Here is an example that shows how to establish this.

Example IAR Injective, Archetype R Archetype R [1779] is the linear transformation

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} -65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5 \\ 36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5 \\ -44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5 \\ 34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5 \\ 12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5 \end{bmatrix}$$

To establish that R is injective we must begin with the assumption that $T(\mathbf{x}) = T(\mathbf{y})$ and somehow arrive from this at the conclusion that $\mathbf{x} = \mathbf{y}$. Here we go,

$$T\left(\mathbf{x}\right) = T\left(\mathbf{y}\right)$$

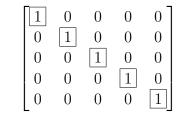
$$T\left(\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}\right) = T\left(\begin{bmatrix}y_1\\y_2\\y_3\\y_4\\y_5\end{bmatrix}\right)$$

$$= T\left(\begin{bmatrix}y_1\\y_2\\y_3\\y_4\\y_5\end{bmatrix}\right)$$

$$\begin{bmatrix}-65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5\\36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5\\-44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5\\34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5\\12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5\end{bmatrix} = \begin{bmatrix}-65y_1 + 128y_2 + 10y_3 - 262y_4 + 40y_5\\36y_1 - 73y_2 - y_3 + 151y_4 - 16y_5\\-44y_1 + 88y_2 + 5y_3 - 180y_4 + 24y_5\\34y_1 - 68y_2 - 3y_3 + 140y_4 - 18y_5\\12y_1 - 24y_2 - y_3 + 49y_4 - 5y_5\end{bmatrix}$$

Now we recognize that we have a homogeneous system of 5 equations in 5 variables (the

terms $x_i - y_i$ are the variables), so we row-reduce the coefficient matrix to

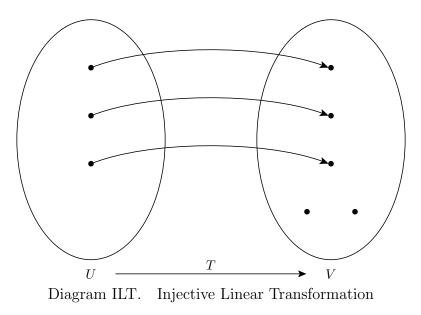


So the only solution is the trivial solution

$$x_1 - y_1 = 0$$
 $x_2 - y_2 = 0$ $x_3 - y_3 = 0$ $x_4 - y_4 = 0$ $x_5 - y_5 = 0$

and we conclude that indeed $\mathbf{x} = \mathbf{y}$. By Definition ILT [1128], T is injective.

Here's the cartoon for an injective linear transformation. It is meant to suggest that we never have two inputs associated with a single output. Again, the two lonely vectors at the bottom of V have no bearing either way on the injectivity of T.



Let's now examine an injective linear transformation between abstract vector spaces.

Example IAV Injective, Archetype V Archetype V [1799] is defined by

$$T: P_3 \to M_{22}, \quad T\left(a+bx+cx^2+dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

To establish that the linear transformation is injective, begin by supposing that two polynomial inputs yield the same output matrix,

$$T(a_1 + b_1x + c_1x^2 + d_1x^3) = T(a_2 + b_2x + c_2x^2 + d_2x^3)$$

Then

$$\mathcal{O} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

= $T \left(a_1 + b_1 x + c_1 x^2 + d_1 x^3 \right) - T \left(a_2 + b_2 x + c_2 x^2 + d_2 x^3 \right)$ Hypothesis
= $T \left((a_1 + b_1 x + c_1 x^2 + d_1 x^3) - (a_2 + b_2 x + c_2 x^2 + d_2 x^3) \right)$ Definition LT [1072]
= $T \left((a_1 - a_2) + (b_1 - b_2) x + (c_1 - c_2) x^2 + (d_1 - d_2) x^3 \right)$ Operations in P_3
= $\begin{bmatrix} (a_1 - a_2) + (b_1 - b_2) & (a_1 - a_2) - 2(c_1 - c_2) \\ (d_1 - d_2) & (b_1 - b_2) - (d_1 - d_2) \end{bmatrix}$ Definition of T

This single matrix equality translates to the homogeneous system of equations in the variables $a_i - b_i$,

$$(a_1 - a_2) + (b_1 - b_2) = 0$$

$$(a_1 - a_2) - 2(c_1 - c_2) = 0$$

$$(d_1 - d_2) = 0$$

$$(b_1 - b_2) - (d_1 - d_2) = 0$$

This system of equations can be rewritten as the matrix equation

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} (a_1 - a_2) \\ (b_1 - b_2) \\ (c_1 - c_2) \\ (d_1 - d_2) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Since the coefficient matrix is nonsingular (check this) the only solution is trivial, i.e.

 $a_1 - a_2 = 0$ $b_1 - b_2 = 0$ $c_1 - c_2 = 0$ $d_1 - d_2 = 0$

so that

$$a_1 = a_2$$
 $b_1 = b_2$ $c_1 = c_2$ $d_1 = d_2$

so the two inputs must be equal polynomials. By Definition ILT [1128], T is injective. \square

Subsection KLT Kernel of a Linear Transformation

For a linear transformation $T: U \to V$, the kernel is a subset of the domain U. Informally, it is the set of all inputs that the transformation sends to the zero vector of the codomain.

It will have some natural connections with the null space of a matrix, so we will keep the same notation, and if you think about your objects, then there should be little confusion. Here's the careful definition.

Definition KLT Kernel of a Linear Transformation

Suppose $T: U \to V$ is a linear transformation. Then the **kernel** of T is the set

$$\mathcal{K}(T) = \{ \mathbf{u} \in U | T(\mathbf{u}) = \mathbf{0} \}$$

(This definition contains Notation KLT.)

Notice that the kernel of T is just the preimage of **0**, $T^{-1}(\mathbf{0})$ (Definition PI [1101]). Here's an example.

Example NKAO Nontrivial kernel, Archetype O

Archetype O [1760] is the linear transformation

$$T: \mathbb{C}^{3} \to \mathbb{C}^{5}, \quad T\left(\begin{bmatrix}x_{1}\\x_{2}\\x_{3}\end{bmatrix}\right) = \begin{bmatrix}-x_{1} + x_{2} - 3x_{3}\\-x_{1} + 2x_{2} - 4x_{3}\\x_{1} + x_{2} + x_{3}\\2x_{1} + 3x_{2} + x_{3}\\x_{1} + 2x_{3}\end{bmatrix}$$

 \triangle

To determine the elements of \mathbb{C}^3 in $\mathcal{K}(T)$, find those vectors **u** such that $T(\mathbf{u}) = \mathbf{0}$, that is,

$$T (\mathbf{u}) = \mathbf{0}$$
$$\begin{bmatrix} -u_1 + u_2 - 3u_3 \\ -u_1 + 2u_2 - 4u_3 \\ u_1 + u_2 + u_3 \\ 2u_1 + 3u_2 + u_3 \\ u_1 + 2u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Vector equality (Definition CVE [199]) leads us to a homogeneous system of 5 equations in the variables u_i ,

$$-u_{1} + u_{2} - 3u_{3} = 0$$

$$-u_{1} + 2u_{2} - 4u_{3} = 0$$

$$u_{1} + u_{2} + u_{3} = 0$$

$$2u_{1} + 3u_{2} + u_{3} = 0$$

$$u_{1} + 2u_{3} = 0$$

Row-reducing the coefficient matrix gives

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The kernel of T is the set of solutions to this homogeneous system of equations, which by Theorem BNS [334] can be expressed as

$$\mathcal{K}(T) = \left\langle \left\{ \begin{bmatrix} -2\\1\\1 \end{bmatrix} \right\} \right\rangle$$

 \boxtimes

We know that the span of a set of vectors is always a subspace (Theorem SSS [709]), so the kernel computed in Example NKAO [1138] is also a subspace. This is no accident, the kernel of a linear transformation is *always* a subspace.

Theorem KLTS

Kernel of a Linear Transformation is a Subspace

Suppose that $T: U \to V$ is a linear transformation. Then the kernel of $T, \mathcal{K}(T)$, is a subspace of U.

Proof We can apply the three-part test of Theorem TSS [698]. First $T(\mathbf{0}_U) = \mathbf{0}_V$ by Theorem LTTZZ [1080], so $\mathbf{0}_U \in \mathcal{K}(T)$ and we know that the kernel is non-empty.

Suppose we assume that $\mathbf{x}, \mathbf{y} \in \mathcal{K}(T)$. Is $\mathbf{x} + \mathbf{y} \in \mathcal{K}(T)$?

$T\left(\mathbf{x}+\mathbf{y}\right) = T\left(\mathbf{x}\right) + T\left(\mathbf{y}\right)$	Definition LT $[1072]$
= 0 + 0	$\mathbf{x}, \mathbf{y} \in \mathcal{K}(T)$
= 0	Property Z $[660]$

This qualifies $\mathbf{x} + \mathbf{y}$ for membership in $\mathcal{K}(T)$. So we have additive closure. Suppose we assume that $\alpha \in \mathbb{C}$ and $\mathbf{x} \in \mathcal{K}(T)$. Is $\alpha \mathbf{x} \in \mathcal{K}(T)$?

$T\left(\alpha\mathbf{x}\right) = \alpha T\left(\mathbf{x}\right)$	Definition LT $[1072]$
$= \alpha 0$	$\mathbf{x} \in \mathcal{K}(T)$
= 0	Theorem ZVSM [677]

This qualifies $\alpha \mathbf{x}$ for membership in $\mathcal{K}(T)$. So we have scalar closure and Theorem TSS [698] tells us that $\mathcal{K}(T)$ is a subspace of U.

Let's compute another kernel, now that we know in advance that it will be a subspace.

Example TKAP Trivial kernel, Archetype P Archetype P [1766] is the linear transformation

$$T: \mathbb{C}^3 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} -x_1 + x_2 + x_3 \\ -x_1 + 2x_2 + 2x_3 \\ x_1 + x_2 + 3x_3 \\ 2x_1 + 3x_2 + x_3 \\ -2x_1 + x_2 + 3x_3 \end{bmatrix}$$

To determine the elements of \mathbb{C}^3 in $\mathcal{K}(T)$, find those vectors **u** such that $T(\mathbf{u}) = \mathbf{0}$, that is,

 $T\left(\mathbf{u}\right)=\mathbf{0}$

$$\begin{bmatrix} -u_1 + u_2 + u_3 \\ -u_1 + 2u_2 + 2u_3 \\ u_1 + u_2 + 3u_3 \\ 2u_1 + 3u_2 + u_3 \\ -2u_1 + u_2 + 3u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Vector equality (Definition CVE [199]) leads us to a homogeneous system of 5 equations in the variables u_i ,

$$-u_1 + u_2 + u_3 = 0$$
$$-u_1 + 2u_2 + 2u_3 = 0$$
$$u_1 + u_2 + 3u_3 = 0$$
$$2u_1 + 3u_2 + u_3 = 0$$
$$-2u_1 + u_2 + 3u_3 = 0$$

Row-reducing the coefficient matrix gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The kernel of T is the set of solutions to this homogeneous system of equations, which is

simply the trivial solution $\mathbf{u} = \mathbf{0}$, so

$$\mathcal{K}(T) = \{\mathbf{0}\} = \langle \{ \} \rangle$$

 \boxtimes

Our next theorem says that if a preimage is a non-empty set then we can construct it by picking any one element and adding on elements of the kernel.

Theorem KPI Kernel and Pre-Image

Suppose $T: U \to V$ is a linear transformation and $\mathbf{v} \in V$. If the preimage $T^{-1}(\mathbf{v})$ is non-empty, and $\mathbf{u} \in T^{-1}(\mathbf{v})$ then

$$T^{-1}(\mathbf{v}) = \{\mathbf{u} + \mathbf{z} | \mathbf{z} \in \mathcal{K}(T)\} = \mathbf{u} + \mathcal{K}(T)$$

Proof Let $M = \{ \mathbf{u} + \mathbf{z} | \mathbf{z} \in \mathcal{K}(T) \}$. First, we show that $M \subseteq T^{-1}(\mathbf{v})$. Suppose that $\mathbf{w} \in M$, so \mathbf{w} has the form $\mathbf{w} = \mathbf{u} + \mathbf{z}$, where $\mathbf{z} \in \mathcal{K}(T)$. Then

$$T (\mathbf{w}) = T (\mathbf{u} + \mathbf{z})$$

= T (\mathbf{u}) + T (\mathbf{z}) Definition LT [1072]
= \mathbf{v} + \mathbf{0} u \in T^{-1} (\mathbf{v}), \mathbf{z} \in \mathcal{K}(T)
= \mathbf{v} Property Z [660]

which qualifies \mathbf{w} for membership in the preimage of $\mathbf{v}, \mathbf{w} \in T^{-1}(\mathbf{v})$. For the opposite inclusion, suppose $\mathbf{x} \in T^{-1}(\mathbf{v})$. Then,

> $T (\mathbf{x} - \mathbf{u}) = T (\mathbf{x}) - T (\mathbf{u})$ = $\mathbf{v} - \mathbf{v}$ = $\mathbf{0}$ Definition LT [1072] $\mathbf{x}, \mathbf{u} \in T^{-1} (\mathbf{v})$

This qualifies $\mathbf{x} - \mathbf{u}$ for membership in the kernel of T, $\mathcal{K}(T)$. So there is a vector $\mathbf{z} \in \mathcal{K}(T)$ such that $\mathbf{x} - \mathbf{u} = \mathbf{z}$. Rearranging this equation gives $\mathbf{x} = \mathbf{u} + \mathbf{z}$ and so $\mathbf{x} \in M$. So $T^{-1}(\mathbf{v}) \subseteq M$ and we see that $M = T^{-1}(\mathbf{v})$, as desired.

This theorem, and its proof, should remind you very much of Theorem PSPHS [256]. Additionally, you might go back and review Example SPIAS [1101]. Can you tell now which is the only preimage to be a subspace?

The next theorem is one we will cite frequently, as it characterizes injections by the size of the kernel.

Theorem KILT

Kernel of an Injective Linear Transformation

Suppose that $T: U \to V$ is a linear transformation. Then T is injective if and only if the kernel of T is trivial, $\mathcal{K}(T) = \{\mathbf{0}\}$.

Proof (\Rightarrow) We assume T is injective and we need to establish that two sets are equal (Definition SE [1604]). Since the kernel is a subspace (Theorem KLTS [1140]), $\{\mathbf{0}\} \subseteq \mathcal{K}(T)$.

To establish the opposite inclusion, suppose $\mathbf{x} \in \mathcal{K}(T)$.

$T\left(\mathbf{x}\right) = 0$	Definition KLT [1138]
$=T\left(0\right)$	Theorem LTTZZ [1080]

We can apply Definition ILT [1128] to conclude that $\mathbf{x} = \mathbf{0}$. Therefore $\mathcal{K}(T) \subseteq \{\mathbf{0}\}$ and by Definition SE [1604], $\mathcal{K}(T) = \{\mathbf{0}\}$.

(\Leftarrow) To establish that T is injective, appeal to Definition ILT [1128] and begin with the assumption that $T(\mathbf{x}) = T(\mathbf{y})$. Then

So $\mathbf{x} - \mathbf{y} \in \mathcal{K}(T)$ by Definition KLT [1138] and with the hypothesis that the kernel is trivial we conclude that $\mathbf{x} - \mathbf{y} = \mathbf{0}$. Then

$$\mathbf{y} = \mathbf{y} + \mathbf{0} = \mathbf{y} + (\mathbf{x} - \mathbf{y}) = \mathbf{x}$$

thus establishing that T is injective by Definition ILT [1128].

Example NIAQR

Not injective, Archetype Q, revisited

We are now in a position to revisit our first example in this section, Example NIAQ [1129]. In that example, we showed that Archetype Q [1771] is not injective by constructing two

vectors, which when used to evaluate the linear transformation provided the same output, thus violating Definition ILT [1128]. Just where did those two vectors come from?

The key is the vector

$$\mathbf{z} = \begin{bmatrix} 3\\4\\1\\3\\3\end{bmatrix}$$

which you can check is an element of $\mathcal{K}(T)$ for Archetype Q [1771]. Choose a vector \mathbf{x} at random, and then compute $\mathbf{y} = \mathbf{x} + \mathbf{z}$ (verify this computation back in Example NIAQ [1129]). Then

$$T (\mathbf{y}) = T (\mathbf{x} + \mathbf{z})$$

= $T (\mathbf{x}) + T (\mathbf{z})$
= $T (\mathbf{x}) + \mathbf{0}$
= $T (\mathbf{x})$
Definition LT [1072]
 $\mathbf{z} \in \mathcal{K}(T)$
Property Z [660]

Whenever the kernel of a linear transformation is non-trivial, we can employ this device and conclude that the linear transformation is not injective. This is another way of viewing Theorem KILT [1144]. For an injective linear transformation, the kernel is trivial and our only choice for \mathbf{z} is the zero vector, which will not help us create two *different* inputs for Tthat yield identical outputs. For every one of the archetypes that is not injective, there is an example presented of exactly this form.

Example NIAO Not injective, Archetype O

In Example NKAO [1138] the kernel of Archetype O [1760] was determined to be

$$\left\langle \left\{ \begin{bmatrix} -2\\1\\1 \end{bmatrix} \right\} \right\rangle$$

a subspace of \mathbb{C}^3 with dimension 1. Since the kernel is not trivial, Theorem KILT [1144] tells us that T is not injective.

Example IAP

Injective, Archetype P

In Example TKAP [1141] it was shown that the linear transformation in Archetype P [1766] has a trivial kernel. So by Theorem KILT [1144], T is injective.

Subsection ILTLI Injective Linear Transformations and Linear Independence

There is a connection between injective linear transformations and linearly independent sets that we will make precise in the next two theorems. However, more informally, we can get a

 \boxtimes

feel for this connection when we think about how each property is defined. A set of vectors is linearly independent if the **only** relation of linear dependence is the trivial one. A linear transformation is injective if the **only** way two input vectors can produce the same output is in the trivial way, when both input vectors are equal.

Theorem ILTLI

Injective Linear Transformations and Linear Independence

Suppose that $T: U \to V$ is an injective linear transformation and $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t\}$ is a linearly independent subset of U. Then $R = \{T(\mathbf{u}_1), T(\mathbf{u}_2), T(\mathbf{u}_3), \ldots, T(\mathbf{u}_t)\}$ is a linearly independent subset of V.

Proof Begin with a relation of linear dependence on R (Definition RLD [735], Definition LI [735]),

$$a_{1}T(\mathbf{u}_{1}) + a_{2}T(\mathbf{u}_{2}) + a_{3}T(\mathbf{u}_{3}) + \dots + a_{t}T(\mathbf{u}_{t}) = \mathbf{0}$$

$$T(a_{1}\mathbf{u}_{1} + a_{2}\mathbf{u}_{2} + a_{3}\mathbf{u}_{3} + \dots + a_{t}\mathbf{u}_{t}) = \mathbf{0}$$
Theorem LTLC [1093]
$$a_{1}\mathbf{u}_{1} + a_{2}\mathbf{u}_{2} + a_{3}\mathbf{u}_{3} + \dots + a_{t}\mathbf{u}_{t} \in \mathcal{K}(T)$$
Definition KLT [1138]
$$a_{1}\mathbf{u}_{1} + a_{2}\mathbf{u}_{2} + a_{3}\mathbf{u}_{3} + \dots + a_{t}\mathbf{u}_{t} \in \{\mathbf{0}\}$$
Theorem KILT [1144]
$$a_{1}\mathbf{u}_{1} + a_{2}\mathbf{u}_{2} + a_{3}\mathbf{u}_{3} + \dots + a_{t}\mathbf{u}_{t} = \mathbf{0}$$
Definition SET [1602]

Since this is a relation of linear dependence on the linearly independent set S, we can conclude that

$$a_1 = 0$$
 $a_2 = 0$ $a_3 = 0$... $a_t = 0$

and this establishes that R is a linearly independent set.

Theorem ILTB Injective Linear Transformations and Bases

Suppose that $T: U \to V$ is a linear transformation and $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_m\}$ is a basis of U. Then T is injective if and only if $C = \{T(\mathbf{u}_1), T(\mathbf{u}_2), T(\mathbf{u}_3), \dots, T(\mathbf{u}_m)\}$ is a linearly independent subset of V.

Proof (\Rightarrow) Assume T is injective. Since B is a basis, we know B is linearly independent (Definition B [777]). Then Theorem ILTLI [1148] says that C is a linearly independent subset of V.

(\Leftarrow) Assume that *C* is linearly independent. To establish that *T* is injective, we will show that the kernel of *T* is trivial (Theorem KILT [1144]). Suppose that $\mathbf{u} \in \mathcal{K}(T)$. As an element of *U*, we can write \mathbf{u} as a linear combination of the basis vectors in *B* (uniquely). So there are scalars, $a_1, a_2, a_3, \ldots, a_m$, such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_m\mathbf{u}_m$$

Then,

$$\mathbf{0} = T (\mathbf{u})$$

$$= T (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_m \mathbf{u}_m)$$

$$= a_1 T (\mathbf{u}_1) + a_2 T (\mathbf{u}_2) + a_3 T (\mathbf{u}_3) + \dots + a_m T (\mathbf{u}_m)$$
Definition KLT [1138]
Definition KLT [1138]
Definition TSVS [746]
Theorem LTLC [1093]

This is a relation of linear dependence (Definition RLD [735]) on the linearly independent set C, so the scalars are all zero: $a_1 = a_2 = a_3 = \cdots = a_m = 0$. Then

$$\mathbf{u} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_m \mathbf{u}_m$$

= $0 \mathbf{u}_1 + 0 \mathbf{u}_2 + 0 \mathbf{u}_3 + \dots + 0 \mathbf{u}_m$ Theorem ZSSM [676]
= $\mathbf{0} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0}$ Theorem ZSSM [676]
= $\mathbf{0}$ Property Z [660]

Since **u** was chosen as an arbitrary vector from $\mathcal{K}(T)$, we have $\mathcal{K}(T) = \{\mathbf{0}\}$ and Theorem KILT [1144] tells us that T is injective.

Subsection ILTD Injective Linear Transformations and Dimension

Theorem ILTD

Injective Linear Transformations and Dimension

Suppose that $T: U \to V$ is an injective linear transformation. Then $\dim(U) \leq \dim(V)$. \Box

Proof Suppose to the contrary that $m = \dim(U) > \dim(V) = t$. Let *B* be a basis of *U*, which will then contain *m* vectors. Apply *T* to each element of *B* to form a set *C* that is a subset of *V*. By Theorem ILTB [1149], *C* is linearly independent and therefore must

contain *m* distinct vectors. So we have found a set of *m* linearly independent vectors in *V*, a vector space of dimension *t*, with m > t. However, this contradicts Theorem G [852], so our assumption is false and dim $(U) \leq \dim(V)$.

Example NIDAU Not injective by dimension, Archetype U The linear transformation in Archetype U [1794] is

$$T: M_{23} \to \mathbb{C}^4, \quad T\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a+2b+12c-3d+e+6f \\ 2a-b-c+d-11f \\ a+b+7c+2d+e-3f \\ a+2b+12c+5e-5f \end{bmatrix}$$

Since dim $(M_{23}) = 6 > 4 = \dim (\mathbb{C}^4)$, T cannot be injective for then T would violate Theorem ILTD [1150].

Notice that the previous example made no use of the actual formula defining the function. Merely a comparison of the dimensions of the domain and codomain are enough to conclude that the linear transformation is not injective. Archetype M [1750] and Archetype N [1755] are two more examples of linear transformations that have "big" domains and "small" codomains, resulting in "collisions" of outputs and thus are non-injective linear transformations.

Subsection CILT Composition of Injective Linear Transformations

In Subsection LT.NLTFO [1105] we saw how to combine linear transformations to build new linear transformations, specifically, how to build the composition of two linear transformations (Definition LTC [1111]). It will be useful later to know that the composition of injective linear transformations is again injective, so we prove that here.

Theorem CILTI

Composition of Injective Linear Transformations is Injective

Suppose that $T: U \to V$ and $S: V \to W$ are injective linear transformations. Then $(S \circ T): U \to W$ is an injective linear transformation.

Proof That the composition is a linear transformation was established in Theorem CLTLT [1111], so we need only establish that the composition is injective. Applying Definition ILT [1128], choose \mathbf{x} , \mathbf{y} from U. Then if $(S \circ T)(\mathbf{x}) = (S \circ T)(\mathbf{y})$,

\Rightarrow	$S\left(T\left(\mathbf{x}\right)\right) = S\left(T\left(\mathbf{y}\right)\right)$	Definition LTC [1111]
\Rightarrow	$T\left(\mathbf{x}\right) = T\left(\mathbf{y}\right)$	Definition ILT [1128] for S
\Rightarrow	$\mathbf{x} = \mathbf{y}$	Definition ILT [1128] for T

Subsection READ Reading Questions

- 1. Suppose $T: \mathbb{C}^8 \to \mathbb{C}^5$ is a linear transformation. Why can't T be injective?
- 2. Describe the kernel of an injective linear transformation.
- 3. Theorem KPI [1143] should remind you of Theorem PSPHS [256]. Why do we say this?

Subsection EXC Exercises

C10 Each archetype below is a linear transformation. Compute the kernel for each. Archetype M [1750] Archetype N [1755] Archetype O [1760] Archetype Q [1771] Archetype Q [1771] Archetype R [1779] Archetype S [1785] Archetype T [1790] Archetype U [1794] Archetype V [1799] Archetype W [1803] Archetype X [1808]

Contributed by Robert Beezer

C20 The linear transformation $T: \mathbb{C}^4 \to \mathbb{C}^3$ is not injective. Find two inputs $\mathbf{x}, \mathbf{y} \in \mathbb{C}^4$

that yield the same output (that is $T(\mathbf{x}) = T(\mathbf{y})$).

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 + x_2 + x_3 \\ -x_1 + 3x_2 + x_3 - x_4 \\ 3x_1 + x_2 + 2x_3 - 2x_4 \end{bmatrix}$$

Contributed by Robert Beezer Solution [1160]

C25 Define the linear transformation

$$T: \mathbb{C}^3 \to \mathbb{C}^2, \quad T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}2x_1 - x_2 + 5x_3\\-4x_1 + 2x_2 - 10x_3\end{bmatrix}$$

Find a basis for the kernel of T, $\mathcal{K}(T)$. Is T injective? Contributed by Robert Beezer Solution [1161]

C26 Let
$$A = \begin{bmatrix} 1 & 2 & 3 & 1 & 0 \\ 2 & -1 & 1 & 0 & 1 \\ 1 & 2 & -1 & -2 & 1 \\ 1 & 3 & 2 & 1 & 2 \end{bmatrix}$$
 and let $T \colon \mathbb{C}^5 \to \mathbb{C}^4$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Is T injective? (Hint: No calculation is required.)

Contributed by Chris Black Solution [1162]

C27 Let
$$T: \mathbb{C}^3 \to \mathbb{C}^3$$
 be given by $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2x+y+z \\ x-y+2z \\ x+2y-z \end{bmatrix}$. Find $\mathcal{K}(T)$. Is T injective? Contributed by Chris Black Solution [1162]

C28 Let
$$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 2 & -1 & 1 & 0 \\ 1 & 2 & -1 & -2 \\ 1 & 3 & 2 & 1 \end{bmatrix}$$
 and let $T \colon \mathbb{C}^4 \to \mathbb{C}^4$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Find $\mathcal{K}(T)$.

Is T injective? Contributed by Chris Black Solution [1163]

C29 Let
$$A = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 2 \\ 1 & 2 & 1 & 1 \end{bmatrix}$$
 and let $T \colon \mathbb{C}^4 \to \mathbb{C}^4$ be given by $T(\mathbf{x}) = A\mathbf{x}$. Find $\mathcal{K}(T)$. Is

T injective?

Contributed by Chris Black Solution [1164]

C30 Let $T: M_{2,2} \to P_2$ be given by $T\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right) = (a+b) + (a+c)x + (a+d)x^2$. Is T injective? Find $\mathcal{K}(T)$. Contributed by Chris Black Solution [1164] **C31** Given that the linear transformation $T: \mathbb{C}^3 \to \mathbb{C}^3$, $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = \begin{bmatrix} 2x+y \\ 2y+z \\ x+2z \end{bmatrix}$ is injective,

show directly that $\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\}$ is a linearly independent set. Contributed by Chris Black Solution [1165]

C32 Given that the linear transformation $T: \mathbb{C}^2 \to \mathbb{C}^3$, $T\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{bmatrix}x+y\\2x+y\\x+2y\end{bmatrix}$ is injective, show directly that $\{T(\mathbf{e}_1), T(\mathbf{e}_2)\}$ is a linearly independent set.

Contributed by Chris Black Solution [1165]

C33 Given that the linear transformation $T: \mathbb{C}^3 \to \mathbb{C}^5$, $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ is

injective, show directly that $\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\}$ is a linearly independent set. Contributed by Chris Black Solution [1166]

C40 Show that the linear transformation R is not injective by finding two different elements of the domain, **x** and **y**, such that $R(\mathbf{x}) = R(\mathbf{y})$. (S_{22} is the vector space of symmetric 2×2

matrices.)

$$R: S_{22} \to P_1 \quad R\left(\begin{bmatrix}a & b\\ b & c\end{bmatrix}\right) = (2a - b + c) + (a + b + 2c)x$$

Contributed by Robert Beezer Solution [1166]

M60 Suppose U and V are vector spaces. Define the function $Z: U \to V$ by $T(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Then by Exercise LT.M60 [1118], Z is a linear transformation. Formulate a condition on U that is equivalent to Z being an injective linear transformation. In other words, fill in the blank to complete the following statement (and then give a proof): Z is injective if and only if U is ______. (See Exercise SLT.M60 [1201], Exercise IVLT.M60 [1248].)

Contributed by Robert Beezer

T10 Suppose $T: U \to V$ is a linear transformation. For which vectors $\mathbf{v} \in V$ is $T^{-1}(\mathbf{v})$ a subspace of U? Contributed by Robert Beezer

T15 Suppose that $T: U \to V$ and $S: V \to W$ are linear transformations. Prove the following relationship between kernels.

$$\mathcal{K}(T) \subseteq \mathcal{K}(S \circ T)$$

Contributed by Robert Beezer Solution [1167]

T20 Suppose that A is an $m \times n$ matrix. Define the linear transformation T by

$$T: \mathbb{C}^n \to \mathbb{C}^m, \quad T(\mathbf{x}) = A\mathbf{x}$$

Prove that the kernel of T equals the null space of A, $\mathcal{K}(T) = \mathcal{N}(A)$. Contributed by Andy Zimmer Solution [1168]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [1154]

A linear transformation that is not injective will have a non-trivial kernel (Theorem KILT [1144]), and this is the key to finding the desired inputs. We need one non-trivial element of the kernel, so suppose that $\mathbf{z} \in \mathbb{C}^4$ is an element of the kernel,

$$\begin{bmatrix} 0\\0\\0 \end{bmatrix} = \mathbf{0} = T(\mathbf{z}) = \begin{bmatrix} 2z_1 + z_2 + z_3\\-z_1 + 3z_2 + z_3 - z_4\\3z_1 + z_2 + 2z_3 - 2z_4 \end{bmatrix}$$

Vector equality Definition CVE [199] leads to the homogeneous system of three equations in four variables,

$$2z_1 + z_2 + z_3 = 0$$

$$-z_1 + 3z_2 + z_3 - z_4 = 0$$

$$3z_1 + z_2 + 2z_3 - 2z_4 = 0$$

The coefficient matrix of this system row-reduces as

$$\begin{bmatrix} 2 & 1 & 1 & 0 \\ -1 & 3 & 1 & -1 \\ 3 & 1 & 2 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \end{bmatrix}$$

From this we can find a solution (we only need one), that is an element of $\mathcal{K}(T)$,

$$\mathbf{z} = \begin{bmatrix} -1\\ -1\\ 3\\ 1 \end{bmatrix}$$

Now, we choose a vector \mathbf{x} at random and set $\mathbf{y} = \mathbf{x} + \mathbf{z}$,

$$\mathbf{x} = \begin{bmatrix} 2\\3\\4\\-2 \end{bmatrix} \qquad \qquad \mathbf{y} = \mathbf{x} + \mathbf{z} = \begin{bmatrix} 2\\3\\4\\-2 \end{bmatrix} + \begin{bmatrix} -1\\-1\\3\\1 \end{bmatrix} = \begin{bmatrix} 1\\2\\7\\-1 \end{bmatrix}$$

and you can check that

$$T\left(\mathbf{x}\right) = \begin{bmatrix} 11\\13\\21 \end{bmatrix} = T\left(\mathbf{y}\right)$$

A quicker solution is to take two elements of the kernel (in this case, scalar multiples of \mathbf{z}) which both get sent to $\mathbf{0}$ by T. Quicker yet, take $\mathbf{0}$ and \mathbf{z} as \mathbf{x} and \mathbf{y} , which also both get sent to $\mathbf{0}$ by T.

C25 Contributed by Robert Beezer Statement [1155] To find the kernel, we require all $\mathbf{x} \in \mathbb{C}^3$ such that $T(\mathbf{x}) = \mathbf{0}$. This condition is

$$\begin{bmatrix} 2x_1 - x_2 + 5x_3\\ -4x_1 + 2x_2 - 10x_3 \end{bmatrix} = \begin{bmatrix} 0\\ 0 \end{bmatrix}$$

This leads to a homogeneous system of two linear equations in three variables, whose coefficient matrix row-reduces to

$$\begin{bmatrix} 1 & -\frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

With two free variables Theorem BNS [334] yields the basis for the null space

$$\left\{ \begin{bmatrix} -\frac{5}{2} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \end{bmatrix} \right\}$$

With $n(T) \neq 0$, $\mathcal{K}(T) \neq \{0\}$, so Theorem KILT [1144] says T is not injective.

C26 Contributed by Chris Black Statement [1155] By Theorem ILTD [1150], if a linear transformation $T: U \to V$ is injective, then $\dim(U) \leq \dim(V)$. In this case, $T: \mathbb{C}^5 \to \mathbb{C}^4$, and $5 = \dim(\mathbb{C}^5) > \dim(\mathbb{C}^4) = 4$. Thus, T cannot possibly be injective.

x - y + 2z = 0

C27 Contributed by Chris Black Statement [1156] If $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \mathbf{0}$, then $\begin{bmatrix} 2x + y + z \\ x - y + 2z \\ x + 2y - z \end{bmatrix} = \mathbf{0}$. Thus, we have the system 2x + y + z = 0

$$x + 2y - z = 0$$

. Thus, we are looking for the nullspace of the matrix $A_T = \begin{bmatrix} 2 & 1 & 1 \\ 1 & -1 & 2 \\ 1 & 2 & -1 \end{bmatrix}$. Since A_T row-

reduces to $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$, the kernel of T is all vectors where x = -z and y = z. Thus, $\mathcal{K}(T) = \left\langle \left\{ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \right\} \right\rangle.$

C28 Contributed by Chris Black Statement [1156] Since T is given by matrix multiplication, $\mathcal{K}(T) = \mathcal{N}(A)$. We have

[1	2	3	1]		[1]		0	0]
2	-1	-	0	RREF	0	1	0	0
1	2	-1	-2	$\rightarrow \rightarrow$		0	1	0 0
[1	3	2	1		0		0	

The nullspace of A is $\{0\}$, so the kernel of T is also trivial: $\mathcal{K}(T) = \{0\}$.

C29 Contributed by Chris Black Statement [1156]

Since T is given by matrix multiplication, $\mathcal{K}(T) = \mathcal{N}(A)$. We have

$$\begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 0 \\ 1 & 2 & 1 & 2 \\ 1 & 2 & 1 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 1/3 & 0 \\ 0 & 1 & 1/3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, a basis for the nullspace of A is $\left\{ \begin{bmatrix} -1 \\ -1 \\ 3 \\ 0 \end{bmatrix} \right\}$, and the kernel is $\mathcal{K}(T) = \left\langle \left\langle \left\langle \begin{bmatrix} -1 \\ -1 \\ 3 \\ 0 \end{bmatrix} \right\rangle \right\rangle$.
Since the kernel is nontrivial, this linear transformation is not injective.

C30 Contributed by Chris Black Statement [1156] We can see without computing that T is not injective, since the degree of $M_{2,2}$ is larger than the degree of P_2 . However, that doesn't address the question of the kernel of T. We need to find all matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ so that $(a + b) + (a + c)x + (a + d)x^2 = 0$. This means a + b = 0, a + c = 0, and a + d = 0, or equivalently, b = d = c = -a. Thus, the kernel is a one-dimensional subspace of $M_{2,2}$ spanned by $\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$. Symbolically, we have

$$\mathcal{K}(T) = \left\langle \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} \right\rangle.$$

C31 Contributed by Chris Black Statement [1157]

We have

$$T\left(\mathbf{e}_{1}\right) = \begin{bmatrix} 2\\0\\1 \end{bmatrix} \qquad \qquad T\left(\mathbf{e}_{2}\right) = \begin{bmatrix} 1\\2\\0 \end{bmatrix} \qquad \qquad T\left(\mathbf{e}_{3}\right) = \begin{bmatrix} 0\\1\\2 \end{bmatrix}$$

Let's put these vectors into a matrix and row reduce to test their linear independence.

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so the set of vectors $\{T(\mathbf{e}_1), T(\mathbf{e}_1), T(\mathbf{e}_1)\}$ is linearly independent.

C32 Contributed by Chris Black Statement [1157] We have $T(\mathbf{e}_1) = \begin{bmatrix} 1\\2\\1 \end{bmatrix}$ and $T(\mathbf{e}_2) = \begin{bmatrix} 1\\1\\2 \end{bmatrix}$. Putting these into a matrix as columns and row-reducing, we have

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 1 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Thus, the set of vectors $\{T(\mathbf{e}_1), T(\mathbf{e}_2)\}$ is linearly independent.

C33 Contributed by Chris Black Statement [1157] We have

$$T(\mathbf{e}_{1}) = \begin{bmatrix} 1\\0\\1\\1\\3 \end{bmatrix} \qquad T(\mathbf{e}_{2}) = \begin{bmatrix} 3\\1\\2\\0\\1 \end{bmatrix} \qquad T(\mathbf{e}_{3}) = \begin{bmatrix} 2\\1\\1\\1\\2 \end{bmatrix}$$

Let's row reduce the matrix of T to test linear independence.

$$\begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 3 & 1 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

so the set of vectors $\{T(\mathbf{e}_1), T(\mathbf{e}_2), T(\mathbf{e}_3)\}$ is linearly independent.

C40 Contributed by Robert Beezer Statement [1157] We choose \mathbf{x} to be any vector we like. A particularly cocky choice would be to choose $\mathbf{x} = \mathbf{0}$, but we will instead choose

$$\mathbf{x} = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$$

Then $R(\mathbf{x}) = 9+9x$. Now compute the kernel of R, which by Theorem KILT [1144] we expect to be nontrivial. Setting $R\left(\begin{bmatrix}a & b\\ b & c\end{bmatrix}\right)$ equal to the zero vector, $\mathbf{0} = 0 + 0x$, and equating coefficients leads to a homogeneous system of equations. Row-reducing the coefficient matrix of this system will allow us to determine the values of a, b and c that create elements of the null space of R,

2	-1	1]	RREF	$\left[1 \right]$	0	1]
1	1	2	/	0	1	1

We only need a single element of the null space of this coefficient matrix, so we will not compute a precise description of the whole null space. Instead, choose the free variable c = 2. Then

$$\mathbf{z} = \begin{bmatrix} -2 & -2\\ -2 & 2 \end{bmatrix}$$

is the corresponding element of the kernel. We compute the desired \mathbf{y} as

$$\mathbf{y} = \mathbf{x} + \mathbf{z} = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix} + \begin{bmatrix} -2 & -2 \\ -2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & -3 \\ -3 & 6 \end{bmatrix}$$

Then check that $R(\mathbf{y}) = 9 + 9x$.

T15 Contributed by Robert Beezer Statement [1158] We are asked to prove that $\mathcal{K}(T)$ is a subset of $\mathcal{K}(S \circ T)$. Employing Definition SSET [1603], choose $\mathbf{x} \in \mathcal{K}(T)$. Then we know that $T(\mathbf{x}) = \mathbf{0}$. So

 $(S \circ T)(\mathbf{x}) = S(T(\mathbf{x}))$ Definition LTC [111]

$$= S(\mathbf{0}) \qquad \mathbf{x} \in \mathcal{K}(T)$$
$$= \mathbf{0} \qquad \text{Theorem LTTZZ [1080]}$$

This qualifies **x** for membership in $\mathcal{K}(S \circ T)$.

T20 Contributed by Andy Zimmer Statement [1159] This is an equality of sets, so we want to establish two subset conditions (Definition SE [1604]).

First, show $\mathcal{N}(A) \subseteq \mathcal{K}(T)$. Choose $\mathbf{x} \in \mathcal{N}(A)$. Check to see if $\mathbf{x} \in \mathcal{K}(T)$,

$$T(\mathbf{x}) = A\mathbf{x}$$
 Definition of T
= **0** $\mathbf{x} \in \mathcal{N}(A)$

So by Definition KLT [1138], $\mathbf{x} \in \mathcal{K}(T)$ and thus $\mathcal{N}(A) \subseteq \mathcal{K}(T)$. Now, show $\mathcal{K}(T) \subseteq \mathcal{N}(A)$. Choose $\mathbf{x} \in \mathcal{K}(T)$. Check to see if $\mathbf{x} \in \mathcal{N}(A)$,

$A\mathbf{x} = T\left(\mathbf{x}\right)$	Definition of T
= 0	$\mathbf{x} \in \mathcal{K}(T)$

So by Definition NSM [152], $\mathbf{x} \in \mathcal{N}(A)$ and thus $\mathcal{K}(T) \subseteq \mathcal{N}(A)$.

Section SLT Surjective Linear Transformations

The companion to an injection is a surjection. Surjective linear transformations are closely related to spanning sets and ranges. So as you read this section reflect back on Section ILT [1128] and note the parallels and the contrasts. In the next section, Section IVLT [1213], we will combine the two properties.

As usual, we lead with a definition.

Definition SLT Surjective Linear Transformation

Suppose $T: U \to V$ is a linear transformation. Then T is **surjective** if for every $\mathbf{v} \in V$ there exists a $\mathbf{u} \in U$ so that $T(\mathbf{u}) = \mathbf{v}$.

Given an arbitrary function, it is possible for there to be an element of the codomain that is not an output of the function (think about the function $y = f(x) = x^2$ and the codomain element y = -3). For a surjective function, this never happens. If we choose any element of the codomain ($\mathbf{v} \in V$) then there must be an input from the domain ($\mathbf{u} \in U$) which will create the output when used to evaluate the linear transformation ($T(\mathbf{u}) = \mathbf{v}$). Some authors prefer the term **onto** where we use surjective, and we will sometimes refer to a surjective linear transformation as a **surjection**.

Subsection ESLT Examples of Surjective Linear Transformations

It is perhaps most instructive to examine a linear transformation that is not surjective first.

Example NSAQ Not surjective, Archetype Q

Archetype Q [1771] is the linear transformation

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} -2x_1 + 3x_2 + 3x_3 - 6x_4 + 3x_5 \\ -16x_1 + 9x_2 + 12x_3 - 28x_4 + 28x_5 \\ -19x_1 + 7x_2 + 14x_3 - 32x_4 + 37x_5 \\ -21x_1 + 9x_2 + 15x_3 - 35x_4 + 39x_5 \\ -9x_1 + 5x_2 + 7x_3 - 16x_4 + 16x_5 \end{bmatrix}$$

We will demonstrate that

$$\mathbf{v} = \begin{bmatrix} -1\\2\\3\\-1\\4 \end{bmatrix}$$

is an unobtainable element of the codomain. Suppose to the contrary that \mathbf{u} is an element

of the domain such that $T(\mathbf{u}) = \mathbf{v}$. Then

$$\begin{bmatrix} -1\\ 2\\ 3\\ -1\\ 4 \end{bmatrix} = \mathbf{v} = T (\mathbf{u}) = T \begin{pmatrix} \begin{bmatrix} u_1\\ u_2\\ u_3\\ u_4\\ u_5 \end{bmatrix} \end{pmatrix}$$
$$= \begin{bmatrix} -2u_1 + 3u_2 + 3u_3 - 6u_4 + 3u_5\\ -16u_1 + 9u_2 + 12u_3 - 28u_4 + 28u_5\\ -19u_1 + 7u_2 + 14u_3 - 32u_4 + 37u_5\\ -21u_1 + 9u_2 + 15u_3 - 35u_4 + 39u_5\\ -9u_1 + 5u_2 + 7u_3 - 16u_4 + 16u_5 \end{bmatrix}$$
$$= \begin{bmatrix} -2 & 3 & 3 & -6 & 3\\ -16 & 9 & 12 & -28 & 28\\ -19 & 7 & 14 & -32 & 37\\ -21 & 9 & 15 & -35 & 39\\ -9 & 5 & 7 & -16 & 16 \end{bmatrix} \begin{bmatrix} u_1\\ u_2\\ u_3\\ u_4\\ u_5 \end{bmatrix}$$

Now we recognize the appropriate input vector \mathbf{u} as a solution to a linear system of equations.

Form the augmented matrix of the system, and row-reduce to

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -\frac{4}{3} & 0 \\ 0 & 0 & 1 & 0 & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

With a leading 1 in the last column, Theorem RCLS [120] tells us the system is inconsistent. From the absence of any solutions we conclude that no such vector \mathbf{u} exists, and by Definition SLT [1169], T is not surjective.

Again, do not concern yourself with how \mathbf{v} was selected, as this will be explained shortly. However, do understand *why* this vector provides enough evidence to conclude that T is not surjective.

To show that a linear transformation is not surjective, it is enough to find a single element of the codomain that is never created by any input, as in Example NSAQ [1170]. However, to show that a linear transformation is surjective we must establish that *every* element of the codomain occurs as an output of the linear transformation for some appropriate input.

Example SAR Surjective, Archetype R Archetype R [1779] is the linear transformation

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} -65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5 \\ 36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5 \\ -44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5 \\ 34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5 \\ 12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5 \end{bmatrix}$$

To establish that R is surjective we must begin with a totally arbitrary element of the codomain, \mathbf{v} and somehow find an input vector \mathbf{u} such that $T(\mathbf{u}) = \mathbf{v}$. We desire,

$$T (\mathbf{u}) = \mathbf{v}$$

$$\begin{bmatrix} -65u_1 + 128u_2 + 10u_3 - 262u_4 + 40u_5\\ 36u_1 - 73u_2 - u_3 + 151u_4 - 16u_5\\ -44u_1 + 88u_2 + 5u_3 - 180u_4 + 24u_5\\ 34u_1 - 68u_2 - 3u_3 + 140u_4 - 18u_5\\ 12u_1 - 24u_2 - u_3 + 49u_4 - 5u_5 \end{bmatrix} = \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\\ v_5 \end{bmatrix}$$

$$\begin{bmatrix} -65 & 128 & 10 & -262 & 40\\ 36 & -73 & -1 & 151 & -16\\ -44 & 88 & 5 & -180 & 24\\ 34 & -68 & -3 & 140 & -18\\ 12 & -24 & -1 & 49 & -5 \end{bmatrix} \begin{bmatrix} u_1\\ u_2\\ u_3\\ u_4\\ u_5 \end{bmatrix} = \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\\ v_5 \end{bmatrix}$$

We recognize this equation as a system of equations in the variables u_i , but our vector of constants contains symbols. In general, we would have to row-reduce the augmented matrix

by hand, due to the symbolic final column. However, in this particular example, the 5×5 coefficient matrix is nonsingular and so has an inverse (Theorem NI [542], Definition MI [506]).

$$\begin{bmatrix} -65 & 128 & 10 & -262 & 40 \\ 36 & -73 & -1 & 151 & -16 \\ -44 & 88 & 5 & -180 & 24 \\ 34 & -68 & -3 & 140 & -18 \\ 12 & -24 & -1 & 49 & -5 \end{bmatrix}^{-1} = \begin{bmatrix} -47 & 92 & 1 & -181 & -14 \\ 27 & -55 & \frac{7}{2} & \frac{221}{2} & 11 \\ -32 & 64 & -1 & -126 & -12 \\ 25 & -50 & \frac{3}{2} & \frac{199}{2} & 9 \\ 9 & -18 & \frac{1}{2} & \frac{71}{2} & 4 \end{bmatrix}$$

so we find that

$$\begin{bmatrix} u_1\\ u_2\\ u_3\\ u_4\\ u_5 \end{bmatrix} = \begin{bmatrix} -47 & 92 & 1 & -181 & -14\\ 27 & -55 & \frac{7}{2} & \frac{221}{2} & 11\\ -32 & 64 & -1 & -126 & -12\\ 25 & -50 & \frac{3}{2} & \frac{199}{2} & 9\\ 9 & -18 & \frac{1}{2} & \frac{71}{2} & 4 \end{bmatrix} \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\\ v_5 \end{bmatrix}$$
$$= \begin{bmatrix} -47v_1 + 92v_2 + v_3 - 181v_4 - 14v_5\\ 27v_1 - 55v_2 + \frac{7}{2}v_3 + \frac{221}{2}v_4 + 11v_5\\ -32v_1 + 64v_2 - v_3 - 126v_4 - 12v_5\\ 25v_1 - 50v_2 + \frac{3}{2}v_3 + \frac{199}{2}v_4 + 9v_5\\ 9v_1 - 18v_2 + \frac{1}{2}v_3 + \frac{71}{2}v_4 + 4v_5 \end{bmatrix}$$

This establishes that if we are given *any* output vector \mathbf{v} , we can use its components in this final expression to formulate a vector \mathbf{u} such that $T(\mathbf{u}) = \mathbf{v}$. So by Definition SLT [1169] we

now know that T is surjective. You might try to verify this condition in its full generality (i.e. evaluate T with this final expression and see if you get \mathbf{v} as the result), or test it more specifically for some numerical vector \mathbf{v} (see Exercise SLT.C20 [1197]).

Let's now examine a surjective linear transformation between abstract vector spaces.

Example SAV Surjective, Archetype V Archetype V [1799] is defined by

$$T: P_3 \to M_{22}, \quad T\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

To establish that the linear transformation is surjective, begin by choosing an arbitrary output. In this example, we need to choose an arbitrary 2×2 matrix, say

$$\mathbf{v} = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$$

and we would like to find an input polynomial

$$\mathbf{u} = a + bx + cx^2 + dx^3$$

so that $T(\mathbf{u}) = \mathbf{v}$. So we have,

$$\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \mathbf{v}$$

Version 2.99

$$= T (\mathbf{u})$$

= $T (a + bx + cx^{2} + dx^{3})$
= $\begin{bmatrix} a + b & a - 2c \\ d & b - d \end{bmatrix}$

Matrix equality leads us to the system of four equations in the four unknowns, x, y, z, w,

$$a + b = x$$
$$a - 2c = y$$
$$d = z$$
$$b - d = w$$

which can be rewritten as a matrix equation,

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

The coefficient matrix is nonsingular, hence it has an inverse,

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 - 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

so we have

$$\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$
$$= \begin{bmatrix} x - z - w \\ z + w \\ \frac{1}{2}(x - y - z - w) \\ z \end{bmatrix}$$

So the input polynomial $\mathbf{u} = (x - z - w) + (z + w)x + \frac{1}{2}(x - y - z - w)x^2 + zx^3$ will yield the output matrix \mathbf{v} , no matter what form \mathbf{v} takes. This means by Definition SLT [1169] that T is surjective. All the same, let's do a concrete demonstration and evaluate T with \mathbf{u} ,

$$T(\mathbf{u}) = T\left((x - z - w) + (z + w)x + \frac{1}{2}(x - y - z - w)x^2 + zx^3 \right)$$

= $\begin{bmatrix} (x - z - w) + (z + w) & (x - z - w) - 2(\frac{1}{2}(x - y - z - w)) \\ z & (z + w) - z \end{bmatrix}$
= $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$
= $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$

 \boxtimes

Subsection RLT Range of a Linear Transformation

For a linear transformation $T: U \to V$, the range is a subset of the codomain V. Informally, it is the set of all outputs that the transformation creates when fed every possible input from the domain. It will have some natural connections with the column space of a matrix, so we will keep the same notation, and if you think about your objects, then there should be little confusion. Here's the careful definition.

Definition RLT Range of a Linear Transformation

Suppose $T: U \to V$ is a linear transformation. Then the **range** of T is the set

 $\mathcal{R}(T) = \{ T(\mathbf{u}) | \mathbf{u} \in U \}$

(This definition contains Notation RLT.)

Example RAO Range, Archetype O \wedge

Archetype O [1760] is the linear transformation

$$T: \mathbb{C}^{3} \to \mathbb{C}^{5}, \quad T\left(\begin{bmatrix} x_{1} \\ x_{2} \\ x_{3} \end{bmatrix} \right) = \begin{bmatrix} -x_{1} + x_{2} - 3x_{3} \\ -x_{1} + 2x_{2} - 4x_{3} \\ x_{1} + x_{2} + x_{3} \\ 2x_{1} + 3x_{2} + x_{3} \\ x_{1} + 2x_{3} \end{bmatrix}$$

To determine the elements of \mathbb{C}^5 in $\mathcal{R}(T)$, find those vectors \mathbf{v} such that $T(\mathbf{u}) = \mathbf{v}$ for some $\mathbf{u} \in \mathbb{C}^3$,

$$\mathbf{v} = T (\mathbf{u})$$

$$= \begin{bmatrix} -u_1 + u_2 - 3u_3 \\ -u_1 + 2u_2 - 4u_3 \\ u_1 + u_2 + u_3 \\ 2u_1 + 3u_2 + u_3 \\ u_1 + 2u_3 \end{bmatrix}$$

$$= \begin{bmatrix} -u_1 \\ -u_1 \\ u_1 \\ 2u_1 \\ u_1 \end{bmatrix} + \begin{bmatrix} u_2 \\ 2u_2 \\ u_2 \\ 3u_2 \\ 0 \end{bmatrix} + \begin{bmatrix} -3u_3 \\ -4u_3 \\ u_3 \\ u_3 \\ u_3 \\ 2u_3 \end{bmatrix}$$

$$= u_1 \begin{bmatrix} -1\\ -1\\ 1\\ 2\\ 1 \end{bmatrix} + u_2 \begin{bmatrix} 1\\ 2\\ 1\\ 3\\ 0 \end{bmatrix} + u_3 \begin{bmatrix} -3\\ -4\\ 1\\ 1\\ 2 \end{bmatrix}$$

This says that every output of $T(\mathbf{v})$ can be written as a linear combination of the three vectors

$$\begin{bmatrix} -1 \\ -1 \\ 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 0 \end{bmatrix} \begin{bmatrix} -3 \\ -4 \\ 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}$$

using the scalars u_1 , u_2 , u_3 . Furthermore, since **u** can be any element of \mathbb{C}^3 , every such linear combination is an output. This means that

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} -1\\-1\\1\\2\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\1\\3\\0 \end{bmatrix}, \begin{bmatrix} -3\\-4\\1\\1\\2 \end{bmatrix} \right\} \right\rangle$$

The three vectors in this spanning set for $\mathcal{R}(T)$ form a linearly dependent set (check this!). So we can find a more economical presentation by any of the various methods from Section CRS [560] and Section FS [609]. We will place the vectors into a matrix as rows, row-reduce, toss out zero rows and appeal to Theorem BRS [583], so we can describe the range of T with a basis,

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} 1\\0\\-3\\-7\\-2 \end{bmatrix}, \begin{bmatrix} 0\\1\\2\\5\\1 \end{bmatrix} \right\} \right\rangle$$

We know that the span of a set of vectors is always a subspace (Theorem SSS [709]), so the range computed in Example RAO [1178] is also a subspace. This is no accident, the range of a linear transformation is *always* a subspace.

Theorem RLTS

Range of a Linear Transformation is a Subspace

Suppose that $T: U \to V$ is a linear transformation. Then the range of $T, \mathcal{R}(T)$, is a subspace of V.

Proof We can apply the three-part test of Theorem TSS [698]. First, $\mathbf{0}_U \in U$ and $T(\mathbf{0}_U) = \mathbf{0}_V$ by Theorem LTTZZ [1080], so $\mathbf{0}_V \in \mathcal{R}(T)$ and we know that the range is non-empty.

Suppose we assume that $\mathbf{x}, \mathbf{y} \in \mathcal{R}(T)$. Is $\mathbf{x} + \mathbf{y} \in \mathcal{R}(T)$? If $\mathbf{x}, \mathbf{y} \in \mathcal{R}(T)$ then we know there are vectors $\mathbf{w}, \mathbf{z} \in U$ such that $T(\mathbf{w}) = \mathbf{x}$ and $T(\mathbf{z}) = \mathbf{y}$. Because U is a vector space,

additive closure (Property AC [660]) implies that $\mathbf{w} + \mathbf{z} \in U$. Then

$$T (\mathbf{w} + \mathbf{z}) = T (\mathbf{w}) + T (\mathbf{z})$$

Definition LT [1072]
$$= \mathbf{x} + \mathbf{y}$$

Definition of **w** and **z**

So we have found an input, $\mathbf{w} + \mathbf{z}$, which when fed into T creates $\mathbf{x} + \mathbf{y}$ as an output. This qualifies $\mathbf{x} + \mathbf{y}$ for membership in $\mathcal{R}(T)$. So we have additive closure.

Suppose we assume that $\alpha \in \mathbb{C}$ and $\mathbf{x} \in \mathcal{R}(T)$. Is $\alpha \mathbf{x} \in \mathcal{R}(T)$? If $\mathbf{x} \in \mathcal{R}(T)$, then there is a vector $\mathbf{w} \in U$ such that $T(\mathbf{w}) = \mathbf{x}$. Because U is a vector space, scalar closure implies that $\alpha \mathbf{w} \in U$. Then

$T\left(\alpha\mathbf{w}\right) = \alpha T\left(\mathbf{w}\right)$	Definition LT $[1072]$
$= \alpha \mathbf{x}$	Definition of \mathbf{w}

So we have found an input $(\alpha \mathbf{w})$ which when fed into T creates $\alpha \mathbf{x}$ as an output. This qualifies $\alpha \mathbf{x}$ for membership in $\mathcal{R}(T)$. So we have scalar closure and Theorem TSS [698] tells us that $\mathcal{R}(T)$ is a subspace of V.

Let's compute another range, now that we know in advance that it will be a subspace.

Example FRAN Full range, Archetype N Archetype N [1755] is the linear transformation

$$T: \mathbb{C}^5 \to \mathbb{C}^3, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 + x_2 + 3x_3 - 4x_4 + 5x_5 \\ x_1 - 2x_2 + 3x_3 - 9x_4 + 3x_5 \\ 3x_1 + 4x_3 - 6x_4 + 5x_5 \end{bmatrix}$$

To determine the elements of \mathbb{C}^3 in $\mathcal{R}(T)$, find those vectors \mathbf{v} such that $T(\mathbf{u}) = \mathbf{v}$ for some $\mathbf{u} \in \mathbb{C}^5$,

$$\mathbf{v} = T (\mathbf{u})$$

$$= \begin{bmatrix} 2u_1 + u_2 + 3u_3 - 4u_4 + 5u_5 \\ u_1 - 2u_2 + 3u_3 - 9u_4 + 3u_5 \\ 3u_1 + 4u_3 - 6u_4 + 5u_5 \end{bmatrix}$$

$$= \begin{bmatrix} 2u_1 \\ u_1 \\ 3u_1 \end{bmatrix} + \begin{bmatrix} u_2 \\ -2u_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 3u_3 \\ 3u_3 \\ 4u_3 \end{bmatrix} + \begin{bmatrix} -4u_4 \\ -9u_4 \\ -6u_4 \end{bmatrix} + \begin{bmatrix} 5u_5 \\ 3u_5 \\ 5u_5 \end{bmatrix}$$

$$= u_1 \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + u_2 \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix} + u_3 \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix} + u_4 \begin{bmatrix} -4 \\ -9 \\ -6 \end{bmatrix} + u_5 \begin{bmatrix} 5 \\ 3 \\ 5 \end{bmatrix}$$

This says that every output of $T(\mathbf{v})$ can be written as a linear combination of the five vectors

$$\begin{bmatrix} 2\\1\\3 \end{bmatrix} \qquad \begin{bmatrix} 1\\-2\\0 \end{bmatrix} \qquad \begin{bmatrix} 3\\3\\4 \end{bmatrix} \qquad \begin{bmatrix} -4\\-9\\-6 \end{bmatrix} \qquad \begin{bmatrix} 5\\3\\5 \end{bmatrix}$$

using the scalars u_1 , u_2 , u_3 , u_4 , u_5 . Furthermore, since **u** can be any element of \mathbb{C}^5 , every such linear combination is an output. This means that

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} 2\\1\\3 \end{bmatrix}, \begin{bmatrix} 1\\-2\\0 \end{bmatrix}, \begin{bmatrix} 3\\3\\4 \end{bmatrix}, \begin{bmatrix} -4\\-9\\-6 \end{bmatrix}, \begin{bmatrix} 5\\3\\5 \end{bmatrix} \right\} \right\rangle$$

The five vectors in this spanning set for $\mathcal{R}(T)$ form a linearly dependent set (Theorem MVSLD [328]). So we can find a more economical presentation by any of the various methods from Section CRS [560] and Section FS [609]. We will place the vectors into a matrix as rows, row-reduce, toss out zero rows and appeal to Theorem BRS [583], so we can describe the range of T with a (nice) basis,

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle = \mathbb{C}^3$$

In contrast to injective linear transformations having small (trivial) kernels (Theorem

 \boxtimes

KILT [1144]), surjective linear transformations have large ranges, as indicated in the next theorem.

Theorem RSLT

Range of a Surjective Linear Transformation

Suppose that $T: U \to V$ is a linear transformation. Then T is surjective if and only if the range of T equals the codomain, $\mathcal{R}(T) = V$.

Proof (\Rightarrow) By Definition RLT [1178], we know that $\mathcal{R}(T) \subseteq V$. To establish the reverse inclusion, assume $\mathbf{v} \in V$. Then since T is surjective (Definition SLT [1169]), there exists a vector $\mathbf{u} \in U$ so that $T(\mathbf{u}) = \mathbf{v}$. However, the existence of \mathbf{u} gains \mathbf{v} membership in $\mathcal{R}(T)$, so $V \subseteq \mathcal{R}(T)$. Thus, $\mathcal{R}(T) = V$.

(⇐) To establish that T is surjective, choose $\mathbf{v} \in V$. Since we are assuming that $\mathcal{R}(T) = V$, $\mathbf{v} \in \mathcal{R}(T)$. This says there is a vector $\mathbf{u} \in U$ so that $T(\mathbf{u}) = \mathbf{v}$, i.e. T is surjective.

Example NSAQR

Not surjective, Archetype Q, revisited

We are now in a position to revisit our first example in this section, Example NSAQ [1170]. In that example, we showed that Archetype Q [1771] is not surjective by constructing a vector in the codomain where no element of the domain could be used to evaluate the linear transformation to create the output, thus violating Definition SLT [1169]. Just where did this vector come from?

The short answer is that the vector

$$\mathbf{v} = \begin{bmatrix} -1\\2\\3\\-1\\4 \end{bmatrix}$$

was constructed to lie outside of the range of T. How was this accomplished? First, the range of T is given by

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0\\0\\1 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\\2 \end{bmatrix} \right\} \right\rangle$$

Suppose an element of the range \mathbf{v}^* has its first 4 components equal to -1, 2, 3, -1, in that order. Then to be an element of $\mathcal{R}(T)$, we would have

$$\mathbf{v}^* = (-1) \begin{bmatrix} 1\\0\\0\\0\\1 \end{bmatrix} + (2) \begin{bmatrix} 0\\1\\0\\-1 \end{bmatrix} + (3) \begin{bmatrix} 0\\0\\1\\0\\-1 \end{bmatrix} + (-1) \begin{bmatrix} 0\\0\\0\\1\\2 \end{bmatrix} = \begin{bmatrix} -1\\2\\3\\-1\\-8 \end{bmatrix}$$

So the only vector in the range with these first four components specified, must have -8 in the fifth component. To set the fifth component to any other value (say, 4) will result in a vector (**v** in Example NSAQ [1170]) outside of the range. Any attempt to find an input for T that will produce **v** as an output will be doomed to failure.

Whenever the range of a linear transformation is not the whole codomain, we can employ this device and conclude that the linear transformation is not surjective. This is another way of viewing Theorem RSLT [1185]. For a surjective linear transformation, the range is all of the codomain and there is no choice for a vector \mathbf{v} that lies in V, yet not in the range. For every one of the archetypes that is not surjective, there is an example presented of exactly this form.

Example NSAO Not surjective, Archetype O

In Example RAO [1178] the range of Archetype O [1760] was determined to be

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} 1\\0\\-3\\-7\\-2 \end{bmatrix}, \begin{bmatrix} 0\\1\\2\\5\\1 \end{bmatrix} \right\} \right\rangle$$

a subspace of dimension 2 in \mathbb{C}^5 . Since $\mathcal{R}(T) \neq \mathbb{C}^5$, Theorem RSLT [1185] says T is not surjective.

Example SAN

Surjective, Archetype N

The range of Archetype N [1755] was computed in Example FRAN [1182] to be

$$\mathcal{R}(T) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle$$

Since the basis for this subspace is the set of standard unit vectors for \mathbb{C}^3 (Theorem SUVB [779]), we have $\mathcal{R}(T) = \mathbb{C}^3$ and by Theorem RSLT [1185], T is surjective.

Subsection SSSLT Spanning Sets and Surjective Linear Transformations

Just as injective linear transformations are allied with linear independence (Theorem ILTLI [1148], Theorem ILTB [1149]), surjective linear transformations are allied with spanning sets.

Theorem SSRLT Spanning Set for Range of a Linear Transformation Suppose that $T: U \to V$ is a linear transformation and $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_t}$ spans U. Then

$$R = \{T(\mathbf{u}_1), T(\mathbf{u}_2), T(\mathbf{u}_3), \dots, T(\mathbf{u}_t)\}\$$

spans $\mathcal{R}(T)$.

Proof We need to establish that $\mathcal{R}(T) = \langle R \rangle$, a set equality. First we establish that $\mathcal{R}(T) \subseteq \langle R \rangle$. To this end, choose $\mathbf{v} \in \mathcal{R}(T)$. Then there exists a vector $\mathbf{u} \in U$, such that $T(\mathbf{u}) = \mathbf{v}$ (Definition RLT [1178]). Because S spans U there are scalars, $a_1, a_2, a_3, \ldots, a_t$, such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_t\mathbf{u}_t$$

Then

$$\mathbf{v} = T (\mathbf{u})$$

$$= T (a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_t \mathbf{u}_t)$$

$$= a_1 T (\mathbf{u}_1) + a_2 T (\mathbf{u}_2) + a_3 T (\mathbf{u}_3) + \dots + a_t T (\mathbf{u}_t)$$
Definition RLT [1178]
Definition RLT [1178]
Definition RLT [1178]
Theorem LTLC [1093]

which establishes that $\mathbf{v} \in \langle R \rangle$ (Definition SS [709]). So $\mathcal{R}(T) \subseteq \langle R \rangle$.

To establish the opposite inclusion, choose an element of the span of R, say $\mathbf{v} \in \langle R \rangle$. Then there are scalars $b_1, b_2, b_3, \ldots, b_t$ so that

$$\mathbf{v} = b_1 T (\mathbf{u}_1) + b_2 T (\mathbf{u}_2) + b_3 T (\mathbf{u}_3) + \dots + b_t T (\mathbf{u}_t)$$
Definition SS [709]
= $T (b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2 + b_3 \mathbf{u}_3 + \dots + b_t \mathbf{u}_t)$ Theorem LTLC [1093]

This demonstrates that **v** is an output of the linear transformation T, so $\mathbf{v} \in \mathcal{R}(T)$. Therefore $\langle R \rangle \subseteq \mathcal{R}(T)$, so we have the set equality $\mathcal{R}(T) = \langle R \rangle$ (Definition SE [1604]). In other words,

Π

R spans $\mathcal{R}(T)$ (Definition TSVS [746]).

Theorem SSRLT [1188] provides an easy way to begin the construction of a basis for the range of a linear transformation, since the construction of a spanning set requires simply evaluating the linear transformation on a spanning set of the domain. In practice the best choice for a spanning set of the domain would be as small as possible, in other words, a basis. The resulting spanning set for the codomain may not be linearly independent, so to find a basis for the range might require tossing out redundant vectors from the spanning set. Here's an example.

Example BRLT

A basis for the range of a linear transformation

Define the linear transformation $T: M_{22} \to P_2$ by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a + 2b + 8c + d) + (-3a + 2b + 5d)x + (a + b + 5c)x^{2}$$

A convenient spanning set for M_{22} is the basis

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

So by Theorem SSRLT [1188], a spanning set for $\mathcal{R}(T)$ is

$$R = \left\{ T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right), T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right), T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right), T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) \right\}$$

$$= \{1 - 3x + x^2, 2 + 2x + x^2, 8 + 5x^2, 1 + 5x\}$$

The set R is not linearly independent, so if we desire a basis for $\mathcal{R}(T)$, we need to eliminate some redundant vectors. Two particular relations of linear dependence on R are

$$(-2)(1 - 3x + x^{2}) + (-3)(2 + 2x + x^{2}) + (8 + 5x^{2}) = 0 + 0x + 0x^{2} = \mathbf{0}$$
$$(1 - 3x + x^{2}) + (-1)(2 + 2x + x^{2}) + (1 + 5x) = 0 + 0x + 0x^{2} = \mathbf{0}$$

These, individually, allow us to remove $8 + 5x^2$ and 1 + 5x from R with out destroying the property that R spans $\mathcal{R}(T)$. The two remaining vectors are linearly independent (check this!), so we can write

$$\mathcal{R}(T) = \left\langle \left\{ 1 - 3x + x^2, \, 2 + 2x + x^2 \right\} \right\rangle$$

and see that $\dim (\mathcal{R}(T)) = 2$.

Elements of the range are precisely those elements of the codomain with non-empty preimages.

Theorem RPI Range and Pre-Image

Suppose that $T: U \to V$ is a linear transformation. Then

$$\mathbf{v} \in \mathcal{R}(T)$$
 if and only if $T^{-1}(\mathbf{v}) \neq \emptyset$

Proof (\Rightarrow) If $\mathbf{v} \in \mathcal{R}(T)$, then there is a vector $\mathbf{u} \in U$ such that $T(\mathbf{u}) = \mathbf{v}$. This qualifies \mathbf{u} for membership in $T^{-1}(\mathbf{v})$, and thus the preimage of \mathbf{v} is not empty.

 \boxtimes

(⇐) Suppose the preimage of **v** is not empty, so we can choose a vector $\mathbf{u} \in U$ such that $T(\mathbf{u}) = \mathbf{v}$. Then $\mathbf{v} \in \mathcal{R}(T)$.

Theorem SLTB

Surjective Linear Transformations and Bases

Suppose that $T: U \to V$ is a linear transformation and $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_m\}$ is a basis of U. Then T is surjective if and only if $C = \{T(\mathbf{u}_1), T(\mathbf{u}_2), T(\mathbf{u}_3), \ldots, T(\mathbf{u}_m)\}$ is a spanning set for V.

Proof (\Rightarrow) Assume *T* is surjective. Since *B* is a basis, we know *B* is a spanning set of *U* (Definition B [777]). Then Theorem SSRLT [1188] says that *C* spans $\mathcal{R}(T)$. But the hypothesis that *T* is surjective means $V = \mathcal{R}(T)$ (Theorem RSLT [1185]), so *C* spans *V*.

(\Leftarrow) Assume that C spans V. To establish that T is surjective, we will show that every element of V is an output of T for some input (Definition SLT [1169]). Suppose that $\mathbf{v} \in V$. As an element of V, we can write \mathbf{v} as a linear combination of the spanning set C. So there are are scalars, $b_1, b_2, b_3, \ldots, b_m$, such that

$$\mathbf{v} = b_1 T \left(\mathbf{u}_1 \right) + b_2 T \left(\mathbf{u}_2 \right) + b_3 T \left(\mathbf{u}_3 \right) + \dots + b_m T \left(\mathbf{u}_m \right)$$

Now define the vector $\mathbf{u} \in U$ by

$$\mathbf{u} = b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2 + b_3 \mathbf{u}_3 + \dots + b_m \mathbf{u}_m$$

Then

$$T\left(\mathbf{u}\right) = T\left(b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 + \dots + b_m\mathbf{u}_m\right)$$

$$= b_1 T (\mathbf{u}_1) + b_2 T (\mathbf{u}_2) + b_3 T (\mathbf{u}_3) + \dots + b_m T (\mathbf{u}_m)$$
 Theorem LTLC [1093]
= **v**

So, given any choice of a vector $\mathbf{v} \in V$, we can design an input $\mathbf{u} \in U$ to produce \mathbf{v} as an output of T. Thus, by Definition SLT [1169], T is surjective.

Subsection SLTD Surjective Linear Transformations and Dimension

Theorem SLTD Surjective Linear Transformations and Dimension

Suppose that $T: U \to V$ is a surjective linear transformation. Then $\dim(U) \ge \dim(V)$. \Box

Proof Suppose to the contrary that $m = \dim(U) < \dim(V) = t$. Let *B* be a basis of *U*, which will then contain *m* vectors. Apply *T* to each element of *B* to form a set *C* that is a subset of *V*. By Theorem SLTB [1192], *C* is spanning set of *V* with *m* or fewer vectors. So we have a set of *m* or fewer vectors that span *V*, a vector space of dimension *t*, with m < t. However, this contradicts Theorem G [852], so our assumption is false and $\dim(U) \ge \dim(V)$.

Example NSDAT Not surjective by dimension, Archetype T The linear transformation in Archetype T [1790] is

$$T: P_4 \rightarrow P_5, \quad T(p(x)) = (x-2)p(x)$$

Since dim $(P_4) = 5 < 6 = \dim(P_5)$, T cannot be surjective for then it would violate Theorem SLTD [1193].

Notice that the previous example made no use of the actual formula defining the function. Merely a comparison of the dimensions of the domain and codomain are enough to conclude that the linear transformation is not surjective. Archetype O [1760] and Archetype P [1766] are two more examples of linear transformations that have "small" domains and "big" codomains, resulting in an inability to create all possible outputs and thus they are non-surjective linear transformations.

Subsection CSLT Composition of Surjective Linear Transformations

In Subsection LT.NLTFO [1105] we saw how to combine linear transformations to build new linear transformations, specifically, how to build the composition of two linear transformations (Definition LTC [1111]). It will be useful later to know that the composition of surjective linear transformations is again surjective, so we prove that here.

Theorem CSLTS Composition of Surjective Linear Transformations is Surjective

Suppose that $T: U \to V$ and $S: V \to W$ are surjective linear transformations. Then $(S \circ T): U \to W$ is a surjective linear transformation.

Proof That the composition is a linear transformation was established in Theorem CLTLT [1111], so we need only establish that the composition is surjective. Applying Definition SLT [1169], choose $\mathbf{w} \in W$.

Because S is surjective, there must be a vector $\mathbf{v} \in V$, such that $S(\mathbf{v}) = \mathbf{w}$. With the existence of \mathbf{v} established, that T is surjective guarantees a vector $\mathbf{u} \in U$ such that $T(\mathbf{u}) = \mathbf{v}$. Now,

$(S \circ T) (\mathbf{u}) = S (T (\mathbf{u}))$	Definition LTC [1111]
$=S\left(\mathbf{v} ight)$	Definition of \mathbf{u}
$=\mathbf{w}$	Definition of \mathbf{v}

This establishes that any element of the codomain (\mathbf{w}) can be created by evaluating $S \circ T$ with the right input (\mathbf{u}) . Thus, by Definition SLT [1169], $S \circ T$ is surjective.

Subsection READ Reading Questions

- 1. Suppose $T: \mathbb{C}^5 \to \mathbb{C}^8$ is a linear transformation. Why can't T be surjective?
- 2. What is the relationship between a surjective linear transformation and its range?
- 3. There are many similarities and differences between injective and surjective linear transformations. Compare and contrast these two different types of linear transformations. (This means going well beyond just stating their definitions.)

Subsection EXC Exercises

C10 Each archetype below is a linear transformation. Compute the range for each. Archetype M [1750] Archetype N [1755] Archetype O [1760] Archetype Q [1771] Archetype R [1779] Archetype R [1779] Archetype T [1790] Archetype U [1794] Archetype V [1799] Archetype W [1803] Archetype X [1808]

Contributed by Robert Beezer

C20 Example SAR [1172] concludes with an expression for a vector $\mathbf{u} \in \mathbb{C}^5$ that we believe will create the vector $\mathbf{v} \in \mathbb{C}^5$ when used to evaluate T. That is, $T(\mathbf{u}) = \mathbf{v}$. Verify this assertion by actually evaluating T with \mathbf{u} . If you don't have the patience to push around all

these symbols, try choosing a numerical instance of \mathbf{v} , compute \mathbf{u} , and then compute $T(\mathbf{u})$, which should result in \mathbf{v} .

Contributed by Robert Beezer

C22 The linear transformation $S: \mathbb{C}^4 \to \mathbb{C}^3$ is not surjective. Find an output $\mathbf{w} \in \mathbb{C}^3$ that has an empty pre-image (that is $S^{-1}(\mathbf{w}) = \emptyset$.)

$$S\left(\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}\right) = \begin{bmatrix}2x_1 + x_2 + 3x_3 - 4x_4\\x_1 + 3x_2 + 4x_3 + 3x_4\\-x_1 + 2x_2 + x_3 + 7x_4\end{bmatrix}$$

Contributed by Robert Beezer Solution [1203]

C23 Determine whether or not the following linear transformation $T: \mathbb{C}^5 \to P_3$ is surjective:

$$T\left(\begin{bmatrix}a\\b\\c\\d\\e\end{bmatrix}\right) = a + (b+c)x + (c+d)x^2 + (d+e)x^3$$

Contributed by Chris Black Solution [1205]

C24 Determine whether or not the linear transformation $T: P_3 \to \mathbb{C}^5$ below is surjective:

$$T\left(a+bx+cx^{2}+dx^{3}\right) = \begin{bmatrix} a+b\\b+c\\c+d\\a+c\\b+d \end{bmatrix}.$$

Contributed by Chris Black Solution [1206]

C25 Define the linear transformation

$$T: \mathbb{C}^3 \to \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 - x_2 + 5x_3 \\ -4x_1 + 2x_2 - 10x_3 \end{bmatrix}$$

Find a basis for the range of T, $\mathcal{R}(T)$. Is T surjective? Contributed by Robert Beezer Solution [1206]

C26 Let
$$T: \mathbb{C}^3 \to \mathbb{C}^3$$
 be given by $T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) = \begin{bmatrix} a+b+2c \\ 2c \\ a+b+c \end{bmatrix}$. Find a basis of $\mathcal{R}(T)$. Is T

surjective?

Contributed by Chris Black Solution [1207]

C27 Let
$$T: \mathbb{C}^3 \to \mathbb{C}^4$$
 be given by $T\left(\begin{bmatrix}a\\b\\c\end{bmatrix}\right) = \begin{bmatrix}a+b-c\\a-b+c\\-a+b+c\\a+b+c\end{bmatrix}$. Find a basis of $\mathcal{R}(T)$. Is T

surjective?

Contributed by Chris Black Solution [1208]

C28 Let
$$T: \mathbb{C}^4 \to M_{2,2}$$
 be given by $T\left(\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \right) = \begin{bmatrix} a+b & a+b+c \\ a+b+c & a+d \end{bmatrix}$. Find a basis of $\mathcal{R}(T)$. Is T surjective?

Contributed by Chris Black Solution [1209]

C29 Let $T: P_2 \to P_4$ be given by $T(p(x)) = x^2 p(x)$. Find a basis of $\mathcal{R}(T)$. Is T surjective? Contributed by Chris Black Solution [1209]

C30 Let $T: P_4 \to P_3$ be given by T(p(x)) = p'(x), where p'(x) is the derivative. Find a basis of $\mathcal{R}(T)$. Is T surjective? Contributed by Chris Black Solution [1210]

C40 Show that the linear transformation T is not surjective by finding an element of the

codomain, \mathbf{v} , such that there is no vector \mathbf{u} with $T(\mathbf{u}) = \mathbf{v}$.

$$T: \mathbb{C}^3 \to \mathbb{C}^3, \quad T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) = \begin{bmatrix} 2a+3b-c \\ 2b-2c \\ a-b+2c \end{bmatrix}$$

Contributed by Robert Beezer Solution [1210]

M60 Suppose U and V are vector spaces. Define the function $Z: U \to V$ by $T(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Then by Exercise LT.M60 [1118], Z is a linear transformation. Formulate a condition on V that is equivalent to Z being an surjective linear transformation. In other words, fill in the blank to complete the following statement (and then give a proof): Z is surjective if and only if V is ______. (See Exercise ILT.M60 [1158], Exercise IVLT.M60 [1248].) Contributed by Robert Beezer

T15 Suppose that $T: U \to V$ and $S: V \to W$ are linear transformations. Prove the following relationship between ranges.

$$\mathcal{R}(S \circ T) \subseteq \mathcal{R}(S)$$

Contributed by Robert Beezer Solution [1211]

T20 Suppose that A is an $m \times n$ matrix. Define the linear transformation T by

$$T \colon \mathbb{C}^n \to \mathbb{C}^m, \quad T(\mathbf{x}) = A\mathbf{x}$$

Prove that the range of T equals the column space of A, $\mathcal{R}(T) = \mathcal{C}(A)$. Contributed by Andy Zimmer Solution [1211]

Subsection SOL Solutions

C22 Contributed by Robert Beezer Statement [1198]

To find an element of \mathbb{C}^3 with an empty pre-image, we will compute the range of the linear transformation $\mathcal{R}(S)$ and then find an element outside of this set.

By Theorem SSRLT [1188] we can evaluate S with the elements of a spanning set of the domain and create a spanning set for the range.

$$S\left(\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\right) = \begin{bmatrix}2\\1\\-1\end{bmatrix} \quad S\left(\begin{bmatrix}0\\1\\0\\0\end{bmatrix}\right) = \begin{bmatrix}1\\3\\2\end{bmatrix} \quad S\left(\begin{bmatrix}0\\0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}3\\4\\1\end{bmatrix} \quad S\left(\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right) = \begin{bmatrix}-4\\3\\7\end{bmatrix}$$
So
$$\mathcal{R}(S) = \left\langle\left\{\begin{bmatrix}2\\1\\-1\end{bmatrix}, \begin{bmatrix}1\\3\\2\end{bmatrix}, \begin{bmatrix}1\\3\\2\end{bmatrix}, \begin{bmatrix}1\\3\\2\end{bmatrix}, \begin{bmatrix}3\\4\\1\end{bmatrix}, \begin{bmatrix}-4\\3\\7\end{bmatrix}\right\}\right\rangle$$

This spanning set is obviously linearly dependent, so we can reduce it to a basis for $\mathcal{R}(S)$ using Theorem BRS [583], where the elements of the spanning set are placed as the rows of

a matrix. The result is that

$$\mathcal{R}(S) = \left\langle \left\{ \begin{bmatrix} 1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\1\\1 \end{bmatrix} \right\} \right\rangle$$

Therefore, the unique vector in $\mathcal{R}(S)$ with a first slot equal to 6 and a second slot equal to 15 will be the linear combination

$$6\begin{bmatrix}1\\0\\-1\end{bmatrix}+15\begin{bmatrix}0\\1\\1\end{bmatrix}=\begin{bmatrix}6\\15\\9\end{bmatrix}$$

So, any vector with first two components equal to 6 and 15, but with a third component different from 9, such as

$$\mathbf{w} = \begin{bmatrix} 6\\15\\-63 \end{bmatrix}$$

will not be an element of the range of S and will therefore have an empty pre-image. Another strategy on this problem is to guess. Almost any vector will lie outside the range of T, you have to be unlucky to randomly choose an element of the range. This is because the codomain has dimension 3, while the range is "much smaller" at a dimension of 2. You still need to check that your guess lies outside of the range, which generally will involve solving a system of equations that turns out to be inconsistent. **C23** Contributed by Chris Black Statement [1198] The linear transformation T is surjective if for any $p(x) = \alpha + \beta x + \gamma x^2 + \delta x^3$, there is a

vector $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix}$ in \mathbb{C}^5 so that $T(\mathbf{u}) = p(x)$. We need to be able to solve the system

 $a = \alpha$ $b + c = \beta$ $c + d = \gamma$ $d + e = \delta$

This system has an infinite number of solutions, one of which is $a = \alpha$, $b = \beta$, c = 0, $d = \gamma$ and $e = \delta - \gamma$, so that

$$T\left(\begin{bmatrix}\alpha\\\beta\\0\\\gamma\\\delta-\gamma\end{bmatrix}\right) = \alpha + (\beta+0)x + (0+\gamma)x^2 + (\gamma+(\delta-\gamma))x^3$$
$$= \alpha + \beta x + \gamma x^2 + \delta x^3$$
$$= p(x).$$

Thus, T is surjective, since for every vector $\mathbf{v} \in P_3$, there exists a vector $\mathbf{u} \in \mathbb{C}^5$ so that $T(\mathbf{u}) = \mathbf{v}$.

C24 Contributed by Chris Black Statement [1199] According to Theorem SLTD [1193], if a linear transformation $T: U \to V$ is surjective, then $\dim(U) \ge \dim(V)$. In this example, $U = P_3$ has dimension 4, and $V = \mathbb{C}^5$ has dimension 5, so T cannot be surjective. (There is no way T can "expand" the domain P_3 to fill the codomain \mathbb{C}^5 .)

C25 Contributed by Robert Beezer Statement [1199] To find the range of T, apply T to the elements of a spanning set for \mathbb{C}^3 as suggested in Theorem SSRLT [1188]. We will use the standard basis vectors (Theorem SUVB [779]).

$$\mathcal{R}(T) = \left\langle \left\{ T\left(\mathbf{e}_{1}\right), T\left(\mathbf{e}_{2}\right), T\left(\mathbf{e}_{3}\right) \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 2\\-4 \end{bmatrix}, \begin{bmatrix} -1\\2 \end{bmatrix}, \begin{bmatrix} 5\\-10 \end{bmatrix} \right\} \right\rangle$$

Each of these vectors is a scalar multiple of the others, so we can toss two of them in reducing the spanning set to a linearly independent set (or be more careful and apply Theorem BCS [569] on a matrix with these three vectors as columns). The result is the basis of the range,

$$\left\{ \begin{bmatrix} 1\\ -2 \end{bmatrix} \right\}$$

With $r(T) \neq 2$, $\mathcal{R}(T) \neq \mathbb{C}^2$, so Theorem RSLT [1185] says T is not surjective.

C26 Contributed by Chris Black Statement [1199] The range of T is

$$\mathcal{R}(T) = \left\{ \begin{bmatrix} a+b+2c\\2c\\a+b+c \end{bmatrix} \middle| a, b, c \in \mathbb{C} \right\}$$
$$= \left\{ a \begin{bmatrix} 1\\0\\1 \end{bmatrix} + b \begin{bmatrix} 1\\0\\1 \end{bmatrix} + c \begin{bmatrix} 2\\2\\1 \end{bmatrix} \middle| a, b, c \in \mathbb{C} \right\}$$
$$= \left\langle \begin{bmatrix} 1\\0\\1 \end{bmatrix}, \begin{bmatrix} 2\\2\\1 \end{bmatrix} \right\rangle$$

Since the vectors $\begin{bmatrix} 1\\0\\1 \end{bmatrix}$ and $\begin{bmatrix} 2\\2\\1 \end{bmatrix}$ are linearly independent (why?), a basis of $\mathcal{R}(T)$ is $\left\{ \begin{bmatrix} 1\\0\\1 \end{bmatrix}, \begin{bmatrix} 2\\2\\1 \end{bmatrix} \right\}$. Since the dimension of the range is 2 and the dimension of the codomain is 3, T is not surjective.

C27 Contributed by Chris Black Statement [1200]

The range of T is

$$\mathcal{R}(T) = \left\{ \begin{bmatrix} a+b-c\\a-b+c\\-a+b+c\\a+b+c \end{bmatrix} \middle| a, b, c \in \mathbb{C} \right\}$$
$$= \left\{ a \begin{bmatrix} 1\\1\\-1\\1\\1 \end{bmatrix} + b \begin{bmatrix} 1\\-1\\1\\1\\1 \end{bmatrix} + c \begin{bmatrix} -1\\1\\1\\1\\1 \end{bmatrix} \middle| a, b, c \in \mathbb{C} \right\}$$
$$= \left\langle \begin{bmatrix} 1\\1\\-1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} 1\\-1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} -1\\1\\1\\1\\1 \end{bmatrix} \right\rangle$$

By row reduction (not shown), we can see that the set

$$\left\{ \begin{bmatrix} 1\\1\\-1\\1 \end{bmatrix}, \begin{bmatrix} 1\\-1\\1\\1 \end{bmatrix}, \begin{bmatrix} -1\\1\\1\\1 \end{bmatrix} \right\}$$

are linearly independent, so is a basis of $\mathcal{R}(T)$. Since the dimension of the range is 3 and the dimension of the codomain is 4, T is not surjective. (We should have anticipated that

T was not surjective since the dimension of the domain is smaller than the dimension of the codomain.)

C28Contributed by Chris BlackStatement [1200]The range of T is

$$\begin{aligned} \mathcal{R}(T) &= \left\{ \begin{bmatrix} a+b & a+b+c \\ a+b+c & a+d \end{bmatrix} \middle| a,b,c,d \in \mathbb{C} \right\} \\ &= \left\{ a \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + b \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \middle| a,b,c,d \in \mathbb{C} \right\} \\ &= \left\langle \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\rangle \\ &= \left\langle \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\rangle. \end{aligned}$$

Can you explain the last equality above?

These three matrices are linearly independent, so a basis of $\mathcal{R}(T)$ is $\left\{ \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$. Thus, T is not surjective, since the range has dimension 3 which is shy of dim $(M_{2,2}) = 4$. (Notice that the range is actually the subspace of symmetric 2×2 matrices in $M_{2,2}$.)

C29 Contributed by Chris Black Statement [1200] If we transform the basis of P_2 , then Theorem SSRLT [1188] guarantees we will have a spanning set of $\mathcal{R}(T)$. A basis of P_2 is $\{1, x, x^2\}$. If we transform the elements of this set, we get the set $\{x^2, x^3, x^4\}$ which is a spanning set for $\mathcal{R}(T)$. These three vectors are linearly independent, so $\{x^2, x^3, x^4\}$ is a basis of $\mathcal{R}(T)$.

C30 Contributed by Chris Black Statement [1200] If we transform the basis of P_4 , then Theorem SSRLT [1188] guarantees we will have a spanning set of $\mathcal{R}(T)$. A basis of P_4 is $\{1, x, x^2, x^3, x^4\}$. If we transform the elements of this set, we get the set $\{0, 1, 2x, 3x^2, 4x^3\}$ which is a spanning set for $\mathcal{R}(T)$. Reducing this to a linearly independent set, we find that $\{1, 2x, 3x^2, 4x^3\}$ is a basis of $\mathcal{R}(T)$. Since $\mathcal{R}(T)$ and P_3 both have dimension 4, T is surjective.

C40 Contributed by Robert Beezer Statement [1200]

We wish to find an output vector \mathbf{v} that has no associated input. This is the same as requiring that there is no solution to the equality

$$\mathbf{v} = T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) = \begin{bmatrix} 2a+3b-c \\ 2b-2c \\ a-b+2c \end{bmatrix} = a \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + b \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} + c \begin{bmatrix} -1 \\ -2 \\ 2 \end{bmatrix}$$

In other words, we would like to find an element of \mathbb{C}^3 not in the set

$$Y = \left\langle \left\{ \begin{bmatrix} 2\\0\\1 \end{bmatrix}, \begin{bmatrix} 3\\2\\-1 \end{bmatrix}, \begin{bmatrix} -1\\-2\\2 \end{bmatrix} \right\} \right\rangle$$

If we make these vectors the rows of a matrix, and row-reduce, Theorem BRS [583] provides

an alternate description of Y,

$$Y = \left\langle \left\{ \begin{bmatrix} 2\\0\\1 \end{bmatrix}, \begin{bmatrix} 0\\4\\-5 \end{bmatrix} \right\} \right\rangle$$

If we add these vectors together, and then change the third component of the result, we will create a vector that lies outside of Y, say $\mathbf{v} = \begin{bmatrix} 2\\4\\9 \end{bmatrix}$.

T15 Contributed by Robert Beezer Statement [1201] This question asks us to establish that one set $(\mathcal{R}(S \circ T))$ is a subset of another $(\mathcal{R}(S))$. Choose an element in the "smaller" set, say $\mathbf{w} \in \mathcal{R}(S \circ T)$. Then we know that there is a vector $\mathbf{u} \in U$ such that

$$\mathbf{w} = (S \circ T) (\mathbf{u}) = S (T (\mathbf{u}))$$

Now define $\mathbf{v} = T(\mathbf{u})$, so that then

$$S\left(\mathbf{v}\right) = S\left(T\left(\mathbf{u}\right)\right) = \mathbf{w}$$

This statement is sufficient to show that $\mathbf{w} \in \mathcal{R}(S)$, so \mathbf{w} is an element of the "larger" set, and $\mathcal{R}(S \circ T) \subseteq \mathcal{R}(S)$.

T20 Contributed by Andy Zimmer Statement [1201] This is an equality of sets, so we want to establish two subset conditions (Definition SE [1604]). First, show $\mathcal{C}(A) \subseteq \mathcal{R}(T)$. Choose $\mathbf{y} \in \mathcal{C}(A)$. Then by Definition CSM [560] and Definition MVP [461] there is a vector $\mathbf{x} \in \mathbb{C}^n$ such that $A\mathbf{x} = \mathbf{y}$. Then

$$T(\mathbf{x}) = A\mathbf{x}$$
 Definition of T
= \mathbf{y}

This statement qualifies **y** as a member of $\mathcal{R}(T)$ (Definition RLT [1178]), so $\mathcal{C}(A) \subseteq \mathcal{R}(T)$.

Now, show $\mathcal{R}(T) \subseteq \mathcal{C}(A)$. Choose $\mathbf{y} \in \mathcal{R}(T)$. Then by Definition RLT [1178], there is a vector \mathbf{x} in \mathbb{C}^n such that $T(\mathbf{x}) = \mathbf{y}$. Then

$$A\mathbf{x} = T(\mathbf{x})$$
 Definition of T
$$= \mathbf{y}$$

So by Definition CSM [560] and Definition MVP [461], y qualifies for membership in $\mathcal{C}(A)$ and so $\mathcal{R}(T) \subseteq \mathcal{C}(A)$.

Section IVLT Invertible Linear Transformations

In this section we will conclude our introduction to linear transformations by bringing together the twin properties of injectivity and surjectivity and consider linear transformations with both of these properties.

Subsection IVLT Invertible Linear Transformations

One preliminary definition, and then we will have our main definition for this section.

Definition IDLT Identity Linear Transformation The identity linear transformation on the vector space W is defined as

$$I_W \colon W \to W, \qquad I_W(\mathbf{w}) = \mathbf{w}$$

 \triangle

Informally, I_W is the "do-nothing" function. You should check that I_W is really a linear

transformation, as claimed, and then compute its kernel and range to see that it is both injective and surjective. All of these facts should be straightforward to verify (Exercise IVLT.T05 [1249]). With this in hand we can make our main definition.

Definition IVLT Invertible Linear Transformations

Suppose that $T\colon U\to V$ is a linear transformation. If there is a function $S\colon V\to U$ such that

$$S \circ T = I_U \qquad \qquad T \circ S = I_V$$

then T is **invertible**. In this case, we call S the **inverse** of T and write $S = T^{-1}$.

Informally, a linear transformation T is invertible if there is a companion linear transformation, S, which "undoes" the action of T. When the two linear transformations are applied consecutively (composition), in either order, the result is to have no real effect. It is entirely analogous to squaring a positive number and then taking its (positive) square root.

Here is an example of a linear transformation that is invertible. As usual at the beginning of a section, do not be concerned with where S came from, just understand how it illustrates Definition IVLT [1214].

Example AIVLT An invertible linear transformation

Archetype V [1799] is the linear transformation

$$T: P_3 \to M_{22}, \quad T\left(a+bx+cx^2+dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

Define the function $S: M_{22} \to P_3$ defined by

$$S\left(\begin{bmatrix}a&b\\c&d\end{bmatrix}\right) = (a-c-d) + (c+d)x + \frac{1}{2}(a-b-c-d)x^2 + cx^3$$

Then

$$(T \circ S) \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = T \left(S \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) \right)$$
$$= T \left((a - c - d) + (c + d)x + \frac{1}{2}(a - b - c - d)x^2 + cx^3 \right)$$
$$= \begin{bmatrix} (a - c - d) + (c + d) & (a - c - d) - 2(\frac{1}{2}(a - b - c - d)) \\ c & (c + d) - c \end{bmatrix}$$
$$= \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
$$= I_{M_{22}} \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right)$$

And

$$(S \circ T) \left(a + bx + cx^{2} + dx^{3} \right) = S \left(T \left(a + bx + cx^{2} + dx^{3} \right) \right)$$
$$= S \left(\begin{bmatrix} a + b & a - 2c \\ d & b - d \end{bmatrix} \right)$$

$$= ((a+b) - d - (b-d)) + (d + (b-d))x + \left(\frac{1}{2}((a+b) - (a-2c) - d - (b-d))\right)x^{2} + (d)x^{3} = a + bx + cx^{2} + dx^{3} = I_{P_{3}}(a + bx + cx^{2} + dx^{3})$$

For now, understand why these computations show that T is invertible, and that $S = T^{-1}$. Maybe even be amazed by how S works so perfectly in concert with T! We will see later just how to arrive at the correct form of S (when it is possible).

It can be as instructive to study a linear transformation that is not invertible.

Example ANILT A non-invertible linear transformation

Consider the linear transformation $T: \mathbb{C}^3 \to M_{22}$ defined by

$$T\left(\begin{bmatrix}a\\b\\c\end{bmatrix}\right) = \begin{bmatrix}a-b & 2a+2b+c\\3a+b+c & -2a-6b-2c\end{bmatrix}$$

Suppose we were to search for an inverse function $S: M_{22} \to \mathbb{C}^3$.

First verify that the 2 × 2 matrix $A = \begin{bmatrix} 5 & 3 \\ 8 & 2 \end{bmatrix}$ is not in the range of T. This will amount

to finding an input to
$$T$$
, $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$, such that

$$a - b = 5$$
$$2a + 2b + c = 3$$
$$3a + b + c = 8$$
$$-2a - 6b - 2c = 2$$

As this system of equations is inconsistent, there is no input column vector, and $A \notin \mathcal{R}(T)$. How should we define S(A)? Note that

$$T(S(A)) = (T \circ S)(A) = I_{M_{22}}(A) = A$$

So any definition we would provide for S(A) must then be a column vector that T sends to A and we would have $A \in \mathcal{R}(T)$, contrary to the definition of T. This is enough to see that there is no function S that will allow us to conclude that T is invertible, since we cannot provide a consistent definition for S(A) if we assume T is invertible.

Even though we now know that T is not invertible, let's not leave this example just yet. Check that

$$T\left(\begin{bmatrix}1\\-2\\4\end{bmatrix}\right) = \begin{bmatrix}3 & 2\\5 & 2\end{bmatrix} = B \qquad T\left(\begin{bmatrix}0\\-3\\8\end{bmatrix}\right) = \begin{bmatrix}3 & 2\\5 & 2\end{bmatrix} = B$$

How would we define S(B)?

$$S(B) = S\left(T\left(\begin{bmatrix}1\\-2\\4\end{bmatrix}\right)\right) = (S \circ T)\left(\begin{bmatrix}1\\-2\\4\end{bmatrix}\right) = I_{\mathbb{C}^3}\left(\begin{bmatrix}1\\-2\\4\end{bmatrix}\right) = \begin{bmatrix}1\\-2\\4\end{bmatrix}$$

or

$$S(B) = S\left(T\left(\begin{bmatrix}0\\-3\\8\end{bmatrix}\right)\right) = (S \circ T)\left(\begin{bmatrix}0\\-3\\8\end{bmatrix}\right) = I_{\mathbb{C}^3}\left(\begin{bmatrix}0\\-3\\8\end{bmatrix}\right) = \begin{bmatrix}0\\-3\\8\end{bmatrix}$$

Which definition should we provide for S(B)? Both are necessary. But then S is not a function. So we have a second reason to know that there is no function S that will allow us to conclude that T is invertible. It happens that there are infinitely many column vectors that S would have to take to B. Construct the kernel of T,

$$\mathcal{K}(T) = \left\langle \left\{ \begin{bmatrix} -1\\ -1\\ 4 \end{bmatrix} \right\} \right\rangle$$

Now choose either of the two inputs used above for T and add to it a scalar multiple of the basis vector for the kernel of T. For example,

$$\mathbf{x} = \begin{bmatrix} 1\\-2\\4 \end{bmatrix} + (-2) \begin{bmatrix} -1\\-1\\4 \end{bmatrix} = \begin{bmatrix} 3\\0\\-4 \end{bmatrix}$$

then verify that $T(\mathbf{x}) = B$. Practice creating a few more inputs for T that would be sent to B, and see why it is hopeless to think that we could ever provide a reasonable definition for S(B)! There is a "whole subspace's worth" of values that S(B) would have to take on. \boxtimes

In Example ANILT [1216] you may have noticed that T is not surjective, since the matrix A was not in the range of T. And T is not injective since there are two different input column vectors that T sends to the matrix B. Linear transformations T that are not surjective lead to putative inverse functions S that are undefined on inputs outside of the range of T. Linear transformations T that are not injective lead to putative inverse functions S that are not injective lead to putative inverse functions S that are not injective lead to putative inverse functions S that are multiply-defined on each of their inputs. We will formalize these ideas in Theorem ILTIS [1221].

But first notice in Definition IVLT [1214] that we only require the inverse (when it exists) to be a function. When it does exist, it too is a linear transformation.

Theorem ILTLT

Inverse of a Linear Transformation is a Linear Transformation

Suppose that $T: U \to V$ is an invertible linear transformation. Then the function $T^{-1}: V \to U$ is a linear transformation. \Box

Proof We work through verifying Definition LT [1072] for T^{-1} , using the fact that T is a linear transformation to obtain the second equality in each half of the proof. To this end, suppose $\mathbf{x}, \mathbf{y} \in V$ and $\alpha \in \mathbb{C}$.

$$T^{-1}(\mathbf{x} + \mathbf{y}) = T^{-1}\left(T\left(T^{-1}(\mathbf{x})\right) + T\left(T^{-1}(\mathbf{y})\right)\right)$$
 Definition IVLT [1214]
= $T^{-1}\left(T\left(T^{-1}(\mathbf{x}) + T^{-1}(\mathbf{y})\right)\right)$ Definition LT [1072]

 $= T^{-1}(\mathbf{x}) + T^{-1}(\mathbf{y})$ Definition IVLT [1214]

Now check the second defining property of a linear transformation for T^{-1} ,

$$T^{-1}(\alpha \mathbf{x}) = T^{-1}(\alpha T(T^{-1}(\mathbf{x})))$$

Definition IVLT [1214]
$$= T^{-1}(T(\alpha T^{-1}(\mathbf{x})))$$

Definition LT [1072]
$$= \alpha T^{-1}(\mathbf{x})$$

Definition IVLT [1214]

So T^{-1} fulfills the requirements of Definition LT [1072] and is therefore a linear transformation. So when T has an inverse, T^{-1} is also a linear transformation. Additionally, T^{-1} is invertible and *its* inverse is what you might expect.

Theorem IILT

Inverse of an Invertible Linear Transformation

Suppose that $T: U \to V$ is an invertible linear transformation. Then T^{-1} is an invertible linear transformation and $(T^{-1})^{-1} = T$.

Proof Because T is invertible, Definition IVLT [1214] tells us there is a function $T^{-1}: V \to U$ such that

$$T^{-1} \circ T = I_U \qquad \qquad T \circ T^{-1} = I_V$$

Additionally, Theorem ILTLT [1219] tells us that T^{-1} is more than just a function, it is a linear transformation. Now view these two statements as properties of the linear transformation T^{-1} . In light of Definition IVLT [1214], they together say that T^{-1} is invertible (let

T play the role of S in the statement of the definition). Furthermore, the inverse of T^{-1} is then T, i.e. $(T^{-1})^{-1} = T$.

Subsection IV Invertibility

We now know what an inverse linear transformation is, but just which linear transformations have inverses? Here is a theorem we have been preparing for all chapter long.

Theorem ILTIS

Invertible Linear Transformations are Injective and Surjective

Suppose $T: U \to V$ is a linear transformation. Then T is invertible if and only if T is injective and surjective.

Proof (\Rightarrow) Since *T* is presumed invertible, we can employ its inverse, T^{-1} (Definition IVLT [1214]). To see that *T* is injective, suppose **x**, **y** $\in U$ and assume that $T(\mathbf{x}) = T(\mathbf{y})$,

 $\mathbf{x} = I_U(\mathbf{x})$ Definition IDLT [1213] $= (T^{-1} \circ T)(\mathbf{x})$ Definition IVLT [1214] $= T^{-1}(T(\mathbf{x}))$ Definition LTC [1111] $= T^{-1}(T(\mathbf{y}))$ Definition ILT [1128]

$$= (T^{-1} \circ T) (\mathbf{y})$$
 Definition LTC [1111]
$$= I_U (\mathbf{y})$$
 Definition IVLT [1214]
$$= \mathbf{y}$$
 Definition IDLT [1213]

So by Definition ILT [1128] T is injective. To check that T is surjective, suppose $\mathbf{v} \in V$. Then $T^{-1}(\mathbf{v})$ is a vector in U. Compute

$$T(T^{-1}(\mathbf{v})) = (T \circ T^{-1})(\mathbf{v}) \qquad \text{Definition LTC [1111]} \\ = I_V(\mathbf{v}) \qquad \text{Definition IVLT [1214]} \\ = \mathbf{v} \qquad \text{Definition IDLT [1213]}$$

So there is an element from U, when used as an input to T (namely $T^{-1}(\mathbf{v})$) that produces the desired output, \mathbf{v} , and hence T is surjective by Definition SLT [1169].

(\Leftarrow) Now assume that T is both injective and surjective. We will build a function $S: V \to U$ that will establish that T is invertible. To this end, choose any $\mathbf{v} \in V$. Since T is surjective, Theorem RSLT [1185] says $\mathcal{R}(T) = V$, so we have $\mathbf{v} \in \mathcal{R}(T)$. Theorem RPI [1191] says that the pre-image of \mathbf{v} , $T^{-1}(\mathbf{v})$, is nonempty. So we can choose a vector from the pre-image of \mathbf{v} , say \mathbf{u} . In other words, there exists $\mathbf{u} \in T^{-1}(\mathbf{v})$.

Since $T^{-1}(\mathbf{v})$ is non-empty, Theorem KPI [1143] then says that

$$T^{-1}(\mathbf{v}) = \{\mathbf{u} + \mathbf{z} | \mathbf{z} \in \mathcal{K}(T)\}$$

However, because T is injective, by Theorem KILT [1144] the kernel is trivial, $\mathcal{K}(T) = \{\mathbf{0}\}$. So the pre-image is a set with just one element, $T^{-1}(\mathbf{v}) = \{\mathbf{u}\}$. Now we can define S by $S(\mathbf{v}) = \mathbf{u}$. This is the key to this half of this proof. Normally the preimage of a vector from the codomain might be an empty set, or an infinite set. But surjectivity requires that the preimage not be empty, and then injectivity limits the preimage to a singleton. Since our choice of \mathbf{v} was arbitrary, we know that every pre-image for T is a set with a single element. This allows us to construct S as a *function*. Now that it is defined, verifying that it is the inverse of T will be easy. Here we go.

Choose $\mathbf{u} \in U$. Define $\mathbf{v} = T(\mathbf{u})$. Then $T^{-1}(\mathbf{v}) = {\mathbf{u}}$, so that $S(\mathbf{v}) = \mathbf{u}$ and,

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u})) = S(\mathbf{v}) = \mathbf{u} = I_U(\mathbf{u})$$

and since our choice of **u** was arbitrary we have function equality, $S \circ T = I_U$.

Now choose $\mathbf{v} \in V$. Define \mathbf{u} to be the single vector in the set $T^{-1}(\mathbf{v})$, in other words, $\mathbf{u} = S(\mathbf{v})$. Then $T(\mathbf{u}) = \mathbf{v}$, so

$$(T \circ S)(\mathbf{v}) = T(S(\mathbf{v})) = T(\mathbf{u}) = \mathbf{v} = I_V(\mathbf{v})$$

and since our choice of **v** was arbitrary we have function equality, $T \circ S = I_V$.

When a linear transformation is both injective and surjective, the pre-image of any element of the codomain is a set of size one (a "singleton"). This fact allowed us to *construct* the inverse linear transformation in one half of the proof of Theorem ILTIS [1221] (see Technique C [1618]). We can follow this approach to construct the inverse of a specific linear transformation, as the next example shows.

Example CIVLT Computing the Inverse of a Linear Transformations

Consider the linear transformation $T: S_{22} \to P_2$ defined by

$$T\left(\begin{bmatrix}a & b\\b & c\end{bmatrix}\right) = (a+b+c) + (-a+2c)x + (2a+3b+6c)x^2$$

T is invertible, which you are able to verify, perhaps by determining that the kernel of T is trivial and the range of T is all of P_2 . This will be easier once we have Theorem RPNDD [1235], which appears later in this section.

By Theorem ILTIS [1221] we know T^{-1} exists, and it will be critical shortly to realize that T^{-1} is automatically known to be a linear transformation as well (Theorem ILTLT [1219]). To determine the complete behavior of $T^{-1}: P_2 \to S_{22}$ we can simply determine its action on a basis for the domain, P_2 . This is the substance of Theorem LTDB [1094], and an excellent example of its application. Choose any basis of P_2 , the simpler the better, such as $B = \{1, x, x^2\}$. Values of T^{-1} for these three basis elements will be the single elements of their preimages. In turn, we have

$$T^{-1}(1)$$
 :

$$T\left(\begin{bmatrix} a & b \\ b & c \end{bmatrix}\right) = 1 + 0x + 0x^{2}$$
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 0 & 2 & 0 \\ 2 & 3 & 6 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & -3 \end{bmatrix}$$

$$\begin{array}{ll} (\text{preimage}) & T^{-1}(1) = \left\{ \begin{bmatrix} -6 & 10 \\ 10 & -3 \end{bmatrix} \right\} \\ (\text{function}) & T^{-1}(1) = \begin{bmatrix} -6 & 10 \\ 10 & -3 \end{bmatrix} \\ T^{-1}(x) : & T\left(\begin{bmatrix} a & b \\ b & c \end{bmatrix} \right) = 0 + 1x + 0x^2 \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & 2 & 1 \\ 2 & 3 & 6 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix} \\ (\text{preimage}) & T^{-1}(x) = \left\{ \begin{bmatrix} -3 & 4 \\ 4 & -1 \end{bmatrix} \right\} \\ (\text{function}) & T^{-1}(x) = \begin{bmatrix} -3 & 4 \\ 4 & -1 \end{bmatrix} \\ T^{-1}(x^2) : & T\left(\begin{bmatrix} a & b \\ b & c \end{bmatrix} \right) = 0 + 0x + 1x^2 \\ \begin{bmatrix} 1 & 1 & 1 & 0 \\ -1 & 0 & 2 & 0 \\ 2 & 3 & 6 & 1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

(preimage)
$$T^{-1}(x^2) = \left\{ \begin{bmatrix} 2 & -3 \\ -3 & 1 \end{bmatrix} \right\}$$

(function)
$$T^{-1}(x^2) = \begin{bmatrix} 2 & -3 \\ -3 & 1 \end{bmatrix}$$

Theorem LTDB [1094] says, informally, "it is enough to know what a linear transformation does to a basis." Formally, we have the outputs of T^{-1} for a basis, so by Theorem LTDB [1094] there is a unique linear transformation with these outputs. So we put this information to work. The key step here is that we can convert any element of P_2 into a linear combination of the elements of the basis B (Theorem VRRB [755]). We are after a "formula" for the value of T^{-1} on a generic element of P_2 , say $p + qx + rx^2$.

$$T^{-1} (p + qx + rx^{2}) = T^{-1} (p(1) + q(x) + r(x^{2}))$$
Theorem VRRB [755]
$$= pT^{-1} (1) + qT^{-1} (x) + rT^{-1} (x^{2})$$
Theorem LTLC [1093]
$$= p \begin{bmatrix} -6 & 10 \\ 10 & -3 \end{bmatrix} + q \begin{bmatrix} -3 & 4 \\ 4 & -1 \end{bmatrix} + r \begin{bmatrix} 2 & -3 \\ -3 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -6p - 3q + 2r & 10p + 4q - 3r \\ 10p + 4q - 3r & -3p - q + r \end{bmatrix}$$

Notice how a linear combination in the domain of T^{-1} has been translated into a linear combination in the codomain of T^{-1} since we know T^{-1} is a linear transformation by Theorem ILTLT [1219].

Also, notice how the augmented matrices used to determine the three pre-images could

be combined into one calculation of a matrix in extended echelon form, reminiscent of a procedure we know for computing the inverse of a matrix (see Example CMI [512]). Hmmmm. \boxtimes

We will make frequent use of the characterization of invertible linear transformations provided by Theorem ILTIS [1221]. The next theorem is a good example of this, and we will use it often, too.

Theorem CIVLT

Composition of Invertible Linear Transformations

Suppose that $T: U \to V$ and $S: V \to W$ are invertible linear transformations. Then the composition, $(S \circ T): U \to W$ is an invertible linear transformation.

Proof Since S and T are both linear transformations, $S \circ T$ is also a linear transformation by Theorem CLTLT [1111]. Since S and T are both invertible, Theorem ILTIS [1221] says that S and T are both injective and surjective. Then Theorem CILTI [1152] says $S \circ T$ is injective, and Theorem CSLTS [1195] says $S \circ T$ is surjective. Now apply the "other half" of Theorem ILTIS [1221] and conclude that $S \circ T$ is invertible.

When a composition is invertible, the inverse is easy to construct.

Theorem ICLT Inverse of a Composition of Linear Transformations Suppose that $T: U \to V$ and $S: V \to W$ are invertible linear transformations. Then $S \circ T$

is invertible and $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$. **Proof** Compute, for all $\mathbf{w} \in W$ $\left((S \circ T) \circ \left(T^{-1} \circ S^{-1} \right) \right) (\mathbf{w}) = S \left(T \left(T^{-1} \left(S^{-1} \left(\mathbf{w} \right) \right) \right) \right)$ $= S\left(I_V\left(S^{-1}\left(\mathbf{w}\right)\right)\right)$ Definition IVLT [1214] $= S\left(S^{-1}\left(\mathbf{w}\right)\right)$ Definition IDLT [1213] Definition IVLT [1214] $= \mathbf{w}$ $=I_{W}(\mathbf{w})$ Definition IDLT [1213] so $(S \circ T) \circ (T^{-1} \circ S^{-1}) = I_W$ and also $((T^{-1} \circ S^{-1}) \circ (S \circ T)) (\mathbf{u}) = T^{-1} (S^{-1} (S (T (\mathbf{u}))))$ $=T^{-1}(I_V(T(\mathbf{u})))$ Definition IVLT [1214] $=T^{-1}(T(\mathbf{u}))$ Definition IDLT [1213] Definition IVLT [1214] $= \mathbf{u}$ $=I_{U}(\mathbf{u})$ Definition IDLT [1213]

so $(T^{-1} \circ S^{-1}) \circ (S \circ T) = I_U$. By Definition IVLT [1214], $S \circ T$ is invertible and $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$.

Notice that this theorem not only establishes what the inverse of $S \circ T$ is, it also duplicates the conclusion of Theorem CIVLT [1227] and also establishes the invertibility of $S \circ T$. But somehow, the proof of Theorem CIVLT [1227] is nicer way to get this property. Does Theorem ICLT [1227] remind you of the flavor of any theorem we have seen about matrices? (Hint: Think about getting dressed.) Hmmmm.

Subsection SI Structure and Isomorphism

A vector space is defined (Definition VS [659]) as a set of objects ("vectors") endowed with a definition of vector addition (+) and a definition of scalar multiplication (written with juxtaposition). Many of our definitions about vector spaces involve linear combinations (Definition LC [707]), such as the span of a set (Definition SS [709]) and linear independence (Definition LI [735]). Other definitions are built up from these ideas, such as bases (Definition B [777]) and dimension (Definition D [815]). The defining properties of a linear transformation require that a function "respect" the operations of the two vector spaces that are the domain and the codomain (Definition LT [1072]). Finally, an invertible linear transformation is one that can be "undone" — it has a companion that reverses its effect. In this subsection we are going to begin to roll all these ideas into one.

A vector space has "structure" derived from definitions of the two operations and the requirement that these operations interact in ways that satisfy the ten properties of Definition VS [659]. When two different vector spaces have an invertible linear transformation defined between them, then we can translate questions about linear combinations (spans, linear independence, bases, dimension) from the first vector space to the second. The answers

obtained in the second vector space can then be translated back, via the inverse linear transformation, and interpreted in the setting of the first vector space. We say that these invertible linear transformations "preserve structure." And we say that the two vector spaces are "structurally the same." The precise term is "isomorphic," from Greek meaning "of the same form." Let's begin to try to understand this important concept.

Definition IVS Isomorphic Vector Spaces

Two vector spaces U and V are **isomorphic** if there exists an invertible linear transformation T with domain U and codomain $V, T: U \to V$. In this case, we write $U \cong V$, and the linear transformation T is known as an **isomorphism** between U and V. \triangle

A few comments on this definition. First, be careful with your language (Technique L [1614]). Two vector spaces are isomorphic, or not. It is a yes/no situation and the term only applies to a pair of vector spaces. Any invertible linear transformation can be called an isomorphism, it is a term that applies to functions. Second, a given pair of vector spaces there might be several different isomorphisms between the two vector spaces. But it only takes the existence of one to call the pair isomorphic. Third, U isomorphic to V, or V isomorphic to U? Doesn't matter, since the inverse linear transformation will provide the needed isomorphism in the "opposite" direction. Being "isomorphic to" is an equivalence relation on the set of all vector spaces (see Theorem SER [1033] for a reminder about equivalence relations).

Example IVSAV Isomorphic vector spaces, Archetype V Archetype V [1799] is a linear transformation from P_3 to M_{22} ,

$$T: P_3 \to M_{22}, \quad T\left(a+bx+cx^2+dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

Since it is injective and surjective, Theorem ILTIS [1221] tells us that it is an invertible linear transformation. By Definition IVS [1230] we say P_3 and M_{22} are isomorphic.

At a basic level, the term "isomorphic" is nothing more than a codeword for the presence of an invertible linear transformation. However, it is also a description of a powerful idea, and this power only becomes apparent in the course of studying examples and related theorems. In this example, we are led to believe that there is nothing "structurally" different about P_3 and M_{22} . In a certain sense they are the same. Not equal, but the same. One is as good as the other. One is just as interesting as the other.

Here is an extremely basic application of this idea. Suppose we want to compute the following linear combination of polynomials in P_3 ,

$$5(2+3x-4x^2+5x^3) + (-3)(3-5x+3x^2+x^3)$$

Rather than doing it straight-away (which is very easy), we will apply the transformation T to convert into a linear combination of matrices, and then compute in M_{22} according to the definitions of the vector space operations there (Example VSM [663]),

$$T\left(5(2+3x-4x^2+5x^3)+(-3)(3-5x+3x^2+x^3)\right)$$

= 5T (2+3x-4x^2+5x^3) + (-3)T (3-5x+3x^2+x^3) Theorem LTLC [1093]

$$= 5 \begin{bmatrix} 5 & 10 \\ 5 & -2 \end{bmatrix} + (-3) \begin{bmatrix} -2 & -3 \\ 1 & -6 \end{bmatrix}$$
 Definition of T
$$= \begin{bmatrix} 31 & 59 \\ 22 & 8 \end{bmatrix}$$
 Operations in M_{22}

Now we will translate our answer back to P_3 by applying T^{-1} , which we found in Example AIVLT [1214],

$$T^{-1}: M_{22} \to P_3, \quad T^{-1}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a - c - d) + (c + d)x + \frac{1}{2}(a - b - c - d)x^2 + cx^3$$

We compute,

$$T^{-1}\left(\begin{bmatrix} 31 & 59\\ 22 & 8 \end{bmatrix}\right) = 1 + 30x - 29x^2 + 22x^3$$

which is, as expected, exactly what we would have computed for the original linear combination had we just used the definitions of the operations in P_3 (Example VSP [665]). Notice this is meant only as an *illustration* and not a suggested route for doing this particular computation.

Checking the dimensions of two vector spaces can be a quick way to establish that they are not isomorphic. Here's the theorem.

Theorem IVSED Isomorphic Vector Spaces have Equal Dimension Suppose U and V are isomorphic vector spaces. Then $\dim(U) = \dim(V)$.

Proof If U and V are isomorphic, there is an invertible linear transformation $T: U \to V$ (Definition IVS [1230]). T is injective by Theorem ILTIS [1221] and so by Theorem ILTD [1150], dim $(U) \leq \dim(V)$. Similarly, T is surjective by Theorem ILTIS [1221] and so by Theorem SLTD [1193], dim $(U) \geq \dim(V)$. The net effect of these two inequalities is that dim $(U) = \dim(V)$.

The contrapositive of Theorem IVSED [1232] says that if U and V have different dimensions, then they are not isomorphic. Dimension is the simplest "structural" characteristic that will allow you to distinguish non-isomorphic vector spaces. For example P_6 is not isomorphic to M_{34} since their dimensions (7 and 12, respectively) are not equal. With tools developed in Section VR [1261] we will be able to establish that the converse of Theorem IVSED [1232] is true. Think about that one for a moment.

Subsection RNLT Rank and Nullity of a Linear Transformation

Just as a matrix has a rank and a nullity, so too do linear transformations. And just like the rank and nullity of a matrix are related (they sum to the number of columns, Theorem RPNC [831]) the rank and nullity of a linear transformation are related. Here are the definitions and theorems, see the Archetypes (Appendix A [1637]) for loads of examples.

Definition ROLT Rank Of a Linear Transformation

Suppose that $T: U \to V$ is a linear transformation. Then the **rank** of T, r(T), is the dimension of the range of T,

 $r(T) = \dim\left(\mathcal{R}(T)\right)$

(This definition contains Notation ROLT.)

Definition NOLT Nullity Of a Linear Transformation

Suppose that $T: U \to V$ is a linear transformation. Then the **nullity** of T, n(T), is the dimension of the kernel of T,

$$n(T) = \dim\left(\mathcal{K}(T)\right)$$

(This definition contains Notation NOLT.)

Here are two quick theorems.

Theorem ROSLT

Rank Of a Surjective Linear Transformation

Suppose that $T: U \to V$ is a linear transformation. Then the rank of T is the dimension of $V, r(T) = \dim(V)$, if and only if T is surjective.

Proof By Theorem RSLT [1185], T is surjective if and only if $\mathcal{R}(T) = V$. Applying

 \triangle

 \triangle

Definition ROLT [1233], $\mathcal{R}(T) = V$ if and only if $r(T) = \dim(\mathcal{R}(T)) = \dim(V)$.

Theorem NOILT

Nullity Of an Injective Linear Transformation

Suppose that $T: U \to V$ is a linear transformation. Then the nullity of T is zero, n(T) = 0, if and only if T is injective.

Proof By Theorem KILT [1144], T is injective if and only if $\mathcal{K}(T) = \{0\}$. Applying Definition NOLT [1234], $\mathcal{K}(T) = \{0\}$ if and only if n(T) = 0.

Just as injectivity and surjectivity come together in invertible linear transformations, there is a clear relationship between rank and nullity of a linear transformation. If one is big, the other is small.

Theorem RPNDD Rank Plus Nullity is Domain Dimension

Suppose that $T: U \to V$ is a linear transformation. Then

$$r(T) + n(T) = \dim(U)$$

Proof Let r = r(T) and s = n(T). Suppose that $R = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_r} \subseteq V$ is a basis of the range of T, $\mathcal{R}(T)$, and $S = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_s} \subseteq U$ is a basis of the kernel of T, $\mathcal{K}(T)$. Note that R and S are possibly empty, which means that some of the sums in this proof are "empty" and are equal to the zero vector.

Because the elements of R are all in the range of T, each must have a non-empty preimage by Theorem RPI [1191]. Choose vectors $\mathbf{w}_i \in U$, $1 \leq i \leq r$ such that $\mathbf{w}_i \in T^{-1}(\mathbf{v}_i)$. So $T(\mathbf{w}_i) = \mathbf{v}_i$, $1 \leq i \leq r$. Consider the set

$$B = \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3, \, \dots, \, \mathbf{u}_s, \, \mathbf{w}_1, \, \mathbf{w}_2, \, \mathbf{w}_3, \, \dots, \, \mathbf{w}_r\}$$

We claim that B is a basis for U.

To establish linear independence for B, begin with a relation of linear dependence on B. So suppose there are scalars $a_1, a_2, a_3, \ldots, a_s$ and $b_1, b_2, b_3, \ldots, b_r$

$$\mathbf{0} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_s\mathbf{u}_s + b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + b_3\mathbf{w}_3 + \dots + b_r\mathbf{w}_r$$

Then

$$\mathbf{0} = T(\mathbf{0})$$
Theorem LTTZZ [1080]

$$= T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_s\mathbf{u}_s + b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + b_3\mathbf{w}_3 + \dots + b_r\mathbf{w}_r)$$
Definition LI [735]

$$= a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) + a_3T(\mathbf{u}_3) + \dots + a_sT(\mathbf{u}_s) + b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \dots + b_rT(\mathbf{w}_r)$$
Theorem LTLC [1093]

$$= a_1\mathbf{0} + a_2\mathbf{0} + a_3\mathbf{0} + \dots + a_s\mathbf{0} + b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \dots + b_rT(\mathbf{w}_r)$$
Definition KLT [1138]

$$= \mathbf{0} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0} + b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \dots + b_rT(\mathbf{w}_r)$$
Theorem ZVSM [677]

$$= b_1 T (\mathbf{w}_1) + b_2 T (\mathbf{w}_2) + b_3 T (\mathbf{w}_3) + \dots + b_r T (\mathbf{w}_r)$$
Property Z [660]
$$= b_1 \mathbf{v}_1 + b_2 \mathbf{v}_2 + b_3 \mathbf{v}_3 + \dots + b_r \mathbf{v}_r$$
Definition PI [1101]

This is a relation of linear dependence on R (Definition RLD [735]), and since R is a linearly independent set (Definition LI [735]), we see that $b_1 = b_2 = b_3 = \ldots = b_r = 0$. Then the original relation of linear dependence on B becomes

$$\begin{aligned}
\mathbf{0} &= a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_s \mathbf{u}_s + 0 \mathbf{w}_1 + 0 \mathbf{w}_2 + \dots + 0 \mathbf{w}_r \\
&= a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_s \mathbf{u}_s + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0} \\
&= a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_s \mathbf{u}_s
\end{aligned}$$
Theorem ZSSM [676]
Property Z [660]

But this is again a relation of linear independence (Definition RLD [735]), now on the set S. Since S is linearly independent (Definition LI [735]), we have $a_1 = a_2 = a_3 = \ldots = a_r = 0$. Since we now know that all the scalars in the relation of linear dependence on B must be zero, we have established the linear independence of S through Definition LI [735].

To now establish that B spans U, choose an arbitrary vector $\mathbf{u} \in U$. Then $T(\mathbf{u}) \in R(T)$, so there are scalars $c_1, c_2, c_3, \ldots, c_r$ such that

$$T\left(\mathbf{u}\right) = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \dots + c_r\mathbf{v}_r$$

Use the scalars $c_1, c_2, c_3, \ldots, c_r$ to define a vector $\mathbf{y} \in U$,

$$\mathbf{y} = c_1 \mathbf{w}_1 + c_2 \mathbf{w}_2 + c_3 \mathbf{w}_3 + \dots + c_r \mathbf{w}_r$$

Then

$$T (\mathbf{u} - \mathbf{y}) = T (\mathbf{u}) - T (\mathbf{y})$$
Theorem LTLC [1093]
$$= T (\mathbf{u}) - T (c_1 \mathbf{w}_1 + c_2 \mathbf{w}_2 + c_3 \mathbf{w}_3 + \dots + c_r \mathbf{w}_r)$$
Substitution
$$= T (\mathbf{u}) - (c_1 T (\mathbf{w}_1) + c_2 T (\mathbf{w}_2) + \dots + c_r T (\mathbf{w}_r))$$
Theorem LTLC [1093]
$$= T (\mathbf{u}) - (c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3 + \dots + c_r \mathbf{v}_r)$$
 $\mathbf{w}_i \in T^{-1} (\mathbf{v}_i)$
$$= T (\mathbf{u}) - T (\mathbf{u})$$
Substitution
$$= \mathbf{0}$$
Property AI [660]

So the vector $\mathbf{u} - \mathbf{y}$ is sent to the zero vector by T and hence is an element of the kernel of T. As such it can be written as a linear combination of the basis vectors for $\mathcal{K}(T)$, the elements of the set S. So there are scalars $d_1, d_2, d_3, \ldots, d_s$ such that

$$\mathbf{u} - \mathbf{y} = d_1 \mathbf{u}_1 + d_2 \mathbf{u}_2 + d_3 \mathbf{u}_3 + \dots + d_s \mathbf{u}_s$$

Then

$$\mathbf{u} = (\mathbf{u} - \mathbf{y}) + \mathbf{y}$$

= $d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + d_3\mathbf{u}_3 + \dots + d_s\mathbf{u}_s + c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + \dots + c_r\mathbf{w}_r$

This says that for any vector, \mathbf{u} , from U, there exist scalars $(d_1, d_2, d_3, \ldots, d_s, c_1, c_2, c_3, \ldots, c_r)$ that form \mathbf{u} as a linear combination of the vectors in the set B. In other words, B spans U (Definition SS [709]).

So B is a basis (Definition B [777]) of U with s + r vectors, and thus

$$\dim\left(U\right) = s + r = n\left(T\right) + r\left(T\right)$$

as desired.

Theorem RPNC [831] said that the rank and nullity of a matrix sum to the number of columns of the matrix. This result is now an easy consequence of Theorem RPNDD [1235] when we consider the linear transformation $T: \mathbb{C}^n \to \mathbb{C}^m$ defined with the $m \times n$ matrix A by $T(\mathbf{x}) = A\mathbf{x}$. The range and kernel of T are identical to the column space and null space of the matrix A (Exercise ILT.T20 [1159], Exercise SLT.T20 [1201]), so the rank and nullity of the matrix A are identical to the rank and nullity of the linear transformation T. The dimension of the domain of T is the dimension of \mathbb{C}^n , exactly the number of columns for the matrix A.

This theorem can be especially useful in determining basic properties of linear transformations. For example, suppose that $T: \mathbb{C}^6 \to \mathbb{C}^6$ is a linear transformation and you are able to quickly establish that the kernel is trivial. Then n(T) = 0. First this means that T is injective by Theorem NOILT [1235]. Also, Theorem RPNDD [1235] becomes

$$6 = \dim (\mathbb{C}^6) = r(T) + n(T) = r(T) + 0 = r(T)$$

So the rank of T is equal to the rank of the codomain, and by Theorem ROSLT [1234] we know T is surjective. Finally, we know T is invertible by Theorem ILTIS [1221]. So from the determination that the kernel is trivial, and consideration of various dimensions, the theorems of this section allow us to conclude the existence of an inverse linear transformation for T.

Similarly, Theorem RPNDD [1235] can be used to provide alternative proofs for Theorem ILTD [1150], Theorem SLTD [1193] and Theorem IVSED [1232]. It would be an interesting exercise to construct these proofs.

It would be instructive to study the archetypes that are linear transformations and see how many of their properties can be deduced just from considering only the dimensions of the domain and codomain. Then add in just knowledge of either the nullity or rank, and so how much more you can learn about the linear transformation. The table preceding all of the archetypes (Appendix A [1637]) could be a good place to start this analysis.

Subsection SLELT Systems of Linear Equations and Linear Transformations

This subsection does not really belong in this section, or any other section, for that matter. It is just the right time to have a discussion about the connections between the central topic of linear algebra, linear transformations, and our motivating topic from Chapter SLE [2], systems of linear equations. We will discuss several theorems we have seen already, but we will also make some forward-looking statements that will be justified in Chapter R [1260].

Archetype D [1670] and Archetype E [1678] are ideal examples to illustrate connections

with linear transformations. Both have the same coefficient matrix,

$$D = \begin{bmatrix} 2 & 1 & 7 & -7 \\ -3 & 4 & -5 & -6 \\ 1 & 1 & 4 & -5 \end{bmatrix}$$

To apply the *theory* of linear transformations to these two archetypes, employ matrix multiplication (Definition MM [468]) and define the linear transformation,

$$T: \mathbb{C}^4 \to \mathbb{C}^3, \quad T(\mathbf{x}) = D\mathbf{x} = x_1 \begin{bmatrix} 2\\-3\\1 \end{bmatrix} + x_2 \begin{bmatrix} 1\\4\\1 \end{bmatrix} + x_3 \begin{bmatrix} 7\\-5\\4 \end{bmatrix} + x_4 \begin{bmatrix} -7\\-6\\-5 \end{bmatrix}$$

Theorem MBLT [1086] tells us that T is indeed a linear transformation. Archetype D [1670] asks for solutions to $\mathcal{LS}(D, \mathbf{b})$, where $\mathbf{b} = \begin{bmatrix} 8\\ -12\\ -4 \end{bmatrix}$. In the language of linear transformations this is equivalent to asking for $T^{-1}(\mathbf{b})$. In the language of vectors and matrices it asks for a linear combination of the four columns of D that will equal \mathbf{b} . One solution listed is $\mathbf{w} = \begin{bmatrix} 7\\ 8\\ 1\\ 3 \end{bmatrix}$. With a non-empty preimage, Theorem KPI [1143] tells us that the complete solution set of the linear system is the preimage of \mathbf{b} ,

$$\mathbf{w} + \mathcal{K}(T) = \{ \mathbf{w} + \mathbf{z} | \, \mathbf{z} \in \mathcal{K}(T) \}$$

The kernel of the linear transformation T is exactly the null space of the matrix D (see Exercise ILT.T20 [1159]), so this approach to the solution set should be reminiscent of Theorem PSPHS [256]. The kernel of the linear transformation is the preimage of the zero vector, exactly equal to the solution set of the homogeneous system $\mathcal{LS}(D, \mathbf{0})$. Since D has a null space of dimension two, every preimage (and in particular the preimage of \mathbf{b}) is as "big" as a subspace of dimension two (but is not a subspace).

Archetype E [1678] is identical to Archetype D [1670] but with a different vector of constants, $\mathbf{d} = \begin{bmatrix} 2\\3\\2 \end{bmatrix}$. We can use the same linear transformation T to discuss this system

of equations since the coefficient matrix is identical. Now the set of solutions to $\mathcal{LS}(D, \mathbf{d})$ is the pre-image of \mathbf{d} , $T^{-1}(\mathbf{d})$. However, the vector \mathbf{d} is not in the range of the linear transformation (nor is it in the column space of the matrix, since these two sets are equal by Exercise SLT.T20 [1201]). So the empty pre-image is equivalent to the inconsistency of the linear system.

These two archetypes each have three equations in four variables, so either the resulting linear systems are inconsistent, or they are consistent and application of Theorem CMVEI [127] tells us that the system has infinitely many solutions. Considering these same parameters for the linear transformation, the dimension of the domain, \mathbb{C}^4 , is four, while the codomain, \mathbb{C}^3 , has dimension three. Then

$$n(T) = \dim (\mathbb{C}^4) - r(T)$$
 Theorem RPNDD [1235]
= 4 - dim ($\mathcal{R}(T)$) Definition ROLT [1233]

$$\geq 4-3 \qquad \qquad \mathcal{R}(T) \text{ subspace of } \mathbb{C}^3$$
$$= 1$$

So the kernel of T is nontrivial simply by considering the dimensions of the domain (number of variables) and the codomain (number of equations). Pre-images of elements of the codomain that are not in the range of T are empty (inconsistent systems). For elements of the codomain that are in the range of T (consistent systems), Theorem KPI [1143] tells us that the pre-images are built from the kernel, and with a non-trivial kernel, these pre-images are infinite (infinitely many solutions).

When do systems of equations have unique solutions? Consider the system of linear equations $\mathcal{LS}(C, \mathbf{f})$ and the linear transformation $S(\mathbf{x}) = C\mathbf{x}$. If S has a trivial kernel, then pre-images will either be empty or be finite sets with single elements. Correspondingly, the coefficient matrix C will have a trivial null space and solution sets will either be empty (inconsistent) or contain a single solution (unique solution). Should the matrix be square and have a trivial null space then we recognize the matrix as being nonsingular. A square matrix means that the corresponding linear transformation, T, has equal-sized domain and codomain. With a nullity of zero, T is injective, and also Theorem RPNDD [1235] tells us that rank of T is equal to the dimension of the domain, which in turn is equal to the dimension of the codomain. In other words, T is surjective. Injective and surjective, and Theorem ILTIS [1221] tells us that T is invertible. Just as we can use the inverse of the coefficient matrix to find the unique solution of any linear system with a nonsingular coefficient matrix (Theorem SNCM [543]), we can use the inverse of the linear transformation to construct the unique element of any pre-image (proof of Theorem ILTIS [1221]).

The executive summary of this discussion is that to every coefficient matrix of a system of linear equations we can associate a natural linear transformation. Solution sets for systems with this coefficient matrix are preimages of elements of the codomain of the linear transformation. For every theorem about systems of linear equations there is an analogue about linear transformations. The theory of linear transformations provides all the tools to recreate the theory of solutions to linear systems of equations.

We will continue this adventure in Chapter R [1260].

Subsection READ Reading Questions

- 1. What conditions allow us to **easily** determine if a linear transformation is invertible?
- 2. What does it mean to say two vector spaces are isomorphic? Both technically, and informally?
- 3. How do linear transformations relate to systems of linear equations?

Subsection EXC Exercises

C10 The archetypes below are linear transformations of the form $T: U \to V$ that are invertible. For each, the inverse linear transformation is given explicitly as part of the archetype's description. Verify for each linear transformation that

$$T^{-1} \circ T = I_U \qquad \qquad T \circ T^{-1} = I_V$$

Archetype R [1779], Archetype V [1799], Archetype W [1803] Contributed by Robert Beezer

C20 Determine if the linear transformation $T: P_2 \to M_{22}$ is (a) injective, (b) surjective, (c) invertible.

$$T(a + bx + cx^{2}) = \begin{bmatrix} a + 2b - 2c & 2a + 2b \\ -a + b - 4c & 3a + 2b + 2c \end{bmatrix}$$

Contributed by Robert Beezer Solution [1251]

C21 Determine if the linear transformation $S: P_3 \to M_{22}$ is (a) injective, (b) surjective,

(c) invertible.

$$S(a + bx + cx^{2} + dx^{3}) = \begin{bmatrix} -a + 4b + c + 2d & 4a - b + 6c - d \\ a + 5b - 2c + 2d & a + 2c + 5d \end{bmatrix}$$

Contributed by Robert Beezer Solution [1252]

C25 For each linear transformation below: (a) Find the matrix representation of T, (b) Calculate n(T), (c) Calculate r(T), (d) Graph the image in either \mathbb{R}^2 or \mathbb{R}^3 as appropriate, (e) How many dimensions are lost?, and (f) How many dimensions are preserved?

1.
$$T: \mathbb{C}^3 \to \mathbb{C}^3$$
 given by $T\left(\begin{bmatrix} x\\ y\\ z \end{bmatrix}\right) = \begin{bmatrix} x\\ x\\ x \end{bmatrix}$
2. $T: \mathbb{C}^3 \to \mathbb{C}^3$ given by $T\left(\begin{bmatrix} x\\ y\\ z \end{bmatrix}\right) = \begin{bmatrix} x\\ y\\ 0 \end{bmatrix}$
3. $T: \mathbb{C}^3 \to \mathbb{C}^2$ given by $T\left(\begin{bmatrix} x\\ y\\ z \end{bmatrix}\right) = \begin{bmatrix} x\\ x \end{bmatrix}$

4.
$$T: \mathbb{C}^3 \to \mathbb{C}^2$$
 given by $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix}$
5. $T: \mathbb{C}^2 \to \mathbb{C}^3$ given by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$
6. $T: \mathbb{C}^2 \to \mathbb{C}^3$ given by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ x+y \end{bmatrix}$

Contributed by Chris Black

C50 Consider the linear transformation $S: M_{12} \to P_1$ from the set of 1×2 matrices to the set of polynomials of degree at most 1, defined by

$$S\left(\begin{bmatrix}a & b\end{bmatrix}\right) = (3a+b) + (5a+2b)x$$

Prove that S is invertible. Then show that the linear transformation

$$R: P_1 \to M_{12}, \quad R(r+sx) = [(2r-s) \quad (-5r+3s)]$$

is the inverse of S, that is $S^{-1} = R$. Contributed by Robert Beezer Solution [1253] **M30** The linear transformation S below is invertible. Find a formula for the inverse linear transformation, S^{-1} .

$$S \colon P_1 \to M_{1,2}, \quad S(a+bx) = \begin{bmatrix} 3a+b & 2a+b \end{bmatrix}$$

Contributed by Robert Beezer Solution [1254]

M31 The linear transformation $R: M_{12} \to M_{21}$ is invertible. Determine a formula for the inverse linear transformation $R^{-1}: M_{21} \to M_{12}$.

$$R\left(\begin{bmatrix}a & b\end{bmatrix}\right) = \begin{bmatrix}a+3b\\4a+11b\end{bmatrix}$$

Contributed by Robert Beezer Solution [1256]

M50 Rework Example CIVLT [1224], only in place of the basis B for P_2 , choose instead to use the basis $C = \{1, 1 + x, 1 + x + x^2\}$. This will complicate writing a generic element of the domain of T^{-1} as a linear combination of the basis elements, and the algebra will be a bit messier, but in the end you should obtain the same formula for T^{-1} . The inverse linear transformation is what it is, and the choice of a particular basis should not influence the outcome.

Contributed by Robert Beezer

M60 Suppose U and V are vector spaces. Define the function $Z: U \to V$ by $T(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Then by Exercise LT.M60 [1118], Z is a linear transformation. Formulate

a condition on U and V that is equivalent to Z being an invertible linear transformation. In other words, fill in the blank to complete the following statement (and then give a proof): Z is invertible if and only if U and V are ______. (See Exercise ILT.M60 [1158], Exercise SLT.M60 [1201], Exercise MR.M60 [1335].) Contributed by Robert Beezer

T05 Prove that the identity linear transformation (Definition IDLT [1213]) is both injective and surjective, and hence invertible. Contributed by Robert Beezer

T15 Suppose that $T: U \to V$ is a surjective linear transformation and dim $(U) = \dim(V)$. Prove that T is injective.

Contributed by Robert Beezer Solution [1257]

T16 Suppose that $T: U \to V$ is an injective linear transformation and dim $(U) = \dim(V)$. Prove that T is surjective. Contributed by Robert Beezer

T30 Suppose that U and V are isomorphic vector spaces. Prove that there are infinitely many isomorphisms between U and V. Contributed by Robert Beezer Solution [1257]

T40 Suppose $T: U \to V$ and $S: V \to W$ are linear transformations and dim $(U) = \dim(V) = \dim(W)$. Suppose that $S \circ T$ is invertible. Prove that S and T are individually invertible (this could be construed as a converse of Theorem CIVLT [1227]).

Contributed by Robert Beezer Solution [1258]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [1245] (a) We will compute the kernel of T. Suppose that $a + bx + cx^2 \in \mathcal{K}(T)$. Then

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = T \left(a + bx + cx^2 \right) = \begin{bmatrix} a + 2b - 2c & 2a + 2b \\ -a + b - 4c & 3a + 2b + 2c \end{bmatrix}$$

and matrix equality (Theorem ME [1016]) yields the homogeneous system of four equations in three variables,

$$a + 2b - 2c = 0$$
$$2a + 2b = 0$$
$$-a + b - 4c = 0$$
$$3a + 2b + 2c = 0$$

The coefficient matrix of this system row-reduces as

$$\begin{bmatrix} 1 & 2 & -2 \\ 2 & 2 & 0 \\ -1 & 1 & -4 \\ 3 & 2 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

From the existence of non-trivial solutions to this system, we can infer non-zero polynomials in $\mathcal{K}(T)$. By Theorem KILT [1144] we then know that T is not injective.

- (b) Since $3 = \dim(P_2) < \dim(M_{22}) = 4$, by Theorem SLTD [1193] T is not surjective.
- (c) Since T is not surjective, it is not invertible by Theorem ILTIS [1221].

C21 Contributed by Robert Beezer Statement [1245] (a) To check injectivity, we compute the kernel of S. To this end, suppose that $a + bx + cx^2 + dx^3 \in \mathcal{K}(S)$, so

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = S\left(a + bx + cx^{2} + dx^{3}\right) = \begin{bmatrix} -a + 4b + c + 2d & 4a - b + 6c - d \\ a + 5b - 2c + 2d & a + 2c + 5d \end{bmatrix}$$

this creates the homogeneous system of four equations in four variables,

$$-a + 4b + c + 2d = 0$$

$$4a - b + 6c - d = 0$$

$$a + 5b - 2c + 2d = 0$$

$$a + 2c + 5d = 0$$

The coefficient matrix of this system row-reduces as,

$$\begin{bmatrix} -1 & 4 & 1 & 2\\ 4 & -1 & 6 & -1\\ 1 & 5 & -2 & 2\\ 1 & 0 & 2 & 5 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We recognize the coefficient matrix as being nonsingular, so the only solution to the system is a = b = c = d = 0, and the kernel of S is trivial, $\mathcal{K}(S) = \{0 + 0x + 0x^2 + 0x^3\}$. By Theorem KILT [1144], we see that S is injective.

(b) We can establish that S is surjective by considering the rank and nullity of S.

$$r(S) = \dim (P_3) - n(S)$$

Theorem RPNDD [1235]
$$= 4 - 0$$

$$= \dim (M_{22})$$

So, $\mathcal{R}(S)$ is a subspace of M_{22} (Theorem RLTS [1181]) whose dimension equals that of M_{22} . By Theorem EDYES [858], we gain the set equality $\mathcal{R}(S) = M_{22}$. Theorem RSLT [1185] then implies that S is surjective.

(c) Since S is both injective and surjective, Theorem ILTIS [1221] says S is invertible.

C50 Contributed by Robert Beezer Statement [1247] Determine the kernel of S first. The condition that $S(\begin{bmatrix} a & b \end{bmatrix}) = \mathbf{0}$ becomes (3a + b) + (5a + 2b)x = 0 + 0x. Equating coefficients of these polynomials yields the system

$$3a + b = 0$$
$$5a + 2b = 0$$

This homogeneous system has a nonsingular coefficient matrix, so the only solution is a = 0, b = 0 and thus

$$\mathcal{K}(S) = \left\{ \begin{bmatrix} 0 & 0 \end{bmatrix} \right\}$$

By Theorem KILT [1144], we know S is injective. With n(S) = 0 we employ Theorem RPNDD [1235] to find

$$r(S) = r(S) + 0 = r(S) + n(S) = \dim(M_{12}) = 2 = \dim(P_1)$$

Since $\mathcal{R}(S) \subseteq P_1$ and dim $(\mathcal{R}(S)) = \dim(P_1)$, we can apply Theorem EDYES [858] to obtain the set equality $\mathcal{R}(S) = P_1$ and therefore S is surjective.

One of the two defining conditions of an invertible linear transformation is (Definition IVLT [1214])

$$(S \circ R) (a + bx) = S (R (a + bx))$$

= $S ([(2a - b) (-5a + 3b)])$
= $(3(2a - b) + (-5a + 3b)) + (5(2a - b) + 2(-5a + 3b)) x$
= $((6a - 3b) + (-5a + 3b)) + ((10a - 5b) + (-10a + 6b)) x$
= $a + bx$
= $I_{P_1} (a + bx)$

That $(R \circ S) (\begin{bmatrix} a & b \end{bmatrix}) = I_{M_{12}} (\begin{bmatrix} a & b \end{bmatrix})$ is similar.

M30 Contributed by Robert Beezer Statement [1248] (Another approach to this solution would follow Example CIVLT [1224].) Suppose that $S^{-1}: M_{1,2} \to P_1$ has a form given by

$$S^{-1} \begin{pmatrix} z & w \end{pmatrix} = (rz + sw) + (pz + qw) x$$

where r, s, p, q are unknown scalars. Then

$$a + bx = S^{-1} (S (a + bx))$$

= $S^{-1} ([3a + b \ 2a + b])$
= $(r(3a + b) + s(2a + b)) + (p(3a + b) + q(2a + b)) x$
= $((3r + 2s)a + (r + s)b) + ((3p + 2q)a + (p + q)b) x$

Equating coefficients of these two polynomials, and then equating coefficients on a and b, gives rise to 4 equations in 4 variables,

$$3r + 2s = 1$$
$$r + s = 0$$
$$3p + 2q = 0$$
$$p + q = 1$$

This system has a unique solution: r = 1, s = -1, p = -2, q = 3. So the desired inverse linear transformation is

$$S^{-1}(z \ w) = (z - w) + (-2z + 3w)x$$

Notice that the system of 4 equations in 4 variables could be split into two systems, each with two equations in two variables (and identical coefficient matrices). After making this

split, the solution might feel like computing the inverse of a matrix (Theorem CINM [516]). Hmmmm.

M31 Contributed by Robert Beezer Statement [1248] (Another approach to this solution would follow Example CIVLT [1224].)

We are given that R is invertible. The inverse linear transformation can be formulated by considering the pre-image of a generic element of the codomain. With injectivity and surjectivity, we know that the pre-image of any element will be a set of size one — it is this lone element that will be the output of the inverse linear transformation.

Suppose that we set $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ as a generic element of the codomain, M_{21} . Then if $\begin{bmatrix} r & s \end{bmatrix} = \mathbf{w} \in R^{-1}(\mathbf{v}),$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{v} = R(\mathbf{w})$$
$$= \begin{bmatrix} r+3s \\ 4r+11s \end{bmatrix}$$

So we obtain the system of two equations in the two variables r and s,

$$r + 3s = x$$
$$4r + 11s = y$$

With a nonsingular coefficient matrix, we can solve the system using the inverse of the

coefficient matrix,

$$r = -11x + 3y$$
$$s = 4x - y$$

So we define,

$$R^{-1}(\mathbf{v}) = R^{-1}\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \mathbf{w} = \begin{bmatrix}r & s\end{bmatrix} = \begin{bmatrix}-11x + 3y & 4x - y\end{bmatrix}$$

T15 Contributed by Robert Beezer Statement [1249]

If T is surjective, then Theorem RSLT [1185] says $\mathcal{R}(T) = V$, so $r(T) = \dim(V)$ by Definition ROLT [1233]. In turn, the hypothesis gives $r(T) = \dim(U)$. Then, using Theorem RPNDD [1235],

$$n(T) = (r(T) + n(T)) - r(T) = \dim(U) - \dim(U) = 0$$

With a null space of zero dimension, $\mathcal{K}(T) = \{0\}$, and by Theorem KILT [1144] we see that T is injective. T is both injective and surjective so by Theorem ILTIS [1221], T is invertible.

T30 Contributed by Robert Beezer Statement [1249] Since U and V are isomorphic, there is at least one isomorphism between them (Definition IVS [1230]), say $T: U \to V$. As such, T is an invertible linear transformation.

For $\alpha \in \mathbb{C}$ define the linear transformation $S: V \to V$ by $S(\mathbf{v}) = \alpha \mathbf{v}$. Convince yourself that when $\alpha \neq 0$, S is an invertible linear transformation (Definition IVLT [1214]). Then

the composition, $S \circ T : U \to V$, is an invertible linear transformation by Theorem CIVLT [1227]. Once convinced that each non-zero value of α gives rise to a different functions for $S \circ T$, then we have constructed infinitely many isomorphisms from U to V.

T40 Contributed by Robert Beezer Statement [1249] Since $S \circ T$ is invertible, by Theorem ILTIS [1221] $S \circ T$ is injective and therefore has a trivial kernel by Theorem KILT [1144]. Then

$$\mathcal{K}(T) \subseteq \mathcal{K}(S \circ T) \qquad \qquad \text{Exercise ILT.T15 [1158]} \\ = \{\mathbf{0}\} \qquad \qquad \text{Theorem KILT [1144]}$$

Since T has a trivial kernel, by Theorem KILT [1144], T is injective. Also,

$$r(T) = \dim(U) - n(T)$$

Theorem RPNDD [1235]
$$= \dim(U) - 0$$

Theorem NOILT [1235]
$$= \dim(V)$$

Hypothesis

Since $\mathcal{R}(T) \subseteq V$, Theorem EDYES [858] gives $\mathcal{R}(T) = V$, so by Theorem RSLT [1185], T is surjective. Finally, by Theorem ILTIS [1221], T is invertible.

Since $S \circ T$ is invertible, by Theorem ILTIS [1221] $S \circ T$ is surjective and therefore has a full range by Theorem RSLT [1185]. Then

$W = \mathcal{R}(S \circ T)$	Theorem RSLT [1185]
$\subseteq \mathcal{R}(S)$	Exercise SLT.T15 [1201]

Since $\mathcal{R}(S) \subseteq W$ we have $\mathcal{R}(S) = W$ and by Theorem RSLT [1185], S is surjective. By an application of Theorem RPNDD [1235] similar to the first part of this solution, we see that S has a trivial kernel, is therefore injective (Theorem KILT [1144]), and thus invertible (Theorem ILTIS [1221]).

Chapter R Representations

Previous work with linear transformations may have convinced you that we can convert most questions about linear transformations into questions about systems of equations or properties of subspaces of \mathbb{C}^m . In this section we begin to make these vague notions precise. We have used the word "representation" prior, but it will get a heavy workout in this chapter. In many ways, everything we have studied so far was in preparation for this chapter.

Section VR Vector Representations

We begin by establishing an invertible linear transformation between any vector space V of dimension m and \mathbb{C}^m . This will allow us to "go back and forth" between the two vector spaces, no matter how abstract the definition of V might be.

Definition VR

Vector Representation

Suppose that V is a vector space with a basis $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n}$. Define a function $\rho_B: V \to \mathbb{C}^n$ as follows. For $\mathbf{w} \in V$ define the column vector $\rho_B(\mathbf{w}) \in \mathbb{C}^n$ by

$$\mathbf{w} = \left[\rho_B\left(\mathbf{w}\right)\right]_1 \mathbf{v}_1 + \left[\rho_B\left(\mathbf{w}\right)\right]_2 \mathbf{v}_2 + \left[\rho_B\left(\mathbf{w}\right)\right]_3 \mathbf{v}_3 + \dots + \left[\rho_B\left(\mathbf{w}\right)\right]_n \mathbf{v}_n$$

(This definition contains Notation VR.)

This definition looks more complicated that it really is, though the form above will be useful in proofs. Simply stated, given $\mathbf{w} \in V$, we write \mathbf{w} as a linear combination of the basis elements of B. It is key to realize that Theorem VRRB [755] guarantees that we can do this for every \mathbf{w} , and furthermore this expression as a linear combination is unique. The resulting scalars are just the entries of the vector $\rho_B(\mathbf{w})$. This discussion should convince

 \wedge

you that ρ_B is "well-defined" as a function. We can determine a precise output for any input. Now we want to establish that ρ_B is a function with additional properties - it is a linear transformation.

Theorem VRLT Vector Representation is a Linear Transformation

The function ρ_B (Definition VR [1261]) is a linear transformation.

Proof We will take a novel approach in this proof. We will construct another function, which we will easily determine is a linear transformation, and then show that this second function is really ρ_B in disguise. Here we go.

Since B is a basis, we can define $T: V \to \mathbb{C}^n$ to be the unique linear transformation such that $T(\mathbf{v}_i) = \mathbf{e}_i, 1 \leq i \leq n$, as guaranteed by Theorem LTDB [1094], and where the \mathbf{e}_i are the standard unit vectors (Definition SUV [409]). Then suppose for an arbitrary $\mathbf{w} \in V$ we have,

$$[T (\mathbf{w})]_{i} = \left[T \left(\sum_{j=1}^{n} \left[\rho_{B} (\mathbf{w}) \right]_{j} \mathbf{v}_{j} \right) \right]_{i}$$
$$= \left[\sum_{j=1}^{n} \left[\rho_{B} (\mathbf{w}) \right]_{j} T (\mathbf{v}_{j}) \right]_{i}$$
$$= \left[\sum_{j=1}^{n} \left[\rho_{B} (\mathbf{w}) \right]_{j} \mathbf{e}_{j} \right]_{i}$$

Definition VR [1261]

Theorem LTLC [1093]

 \square

$$= \sum_{j=1}^{n} \left[\left[\rho_{B} \left(\mathbf{w} \right) \right]_{j} \mathbf{e}_{j} \right]_{i}$$
 Definition CVA [201]
$$= \sum_{j=1}^{n} \left[\rho_{B} \left(\mathbf{w} \right) \right]_{j} \left[\mathbf{e}_{j} \right]_{i}$$
 Definition CVSM [203]
$$= \left[\rho_{B} \left(\mathbf{w} \right) \right]_{i} \left[\mathbf{e}_{i} \right]_{i} + \sum_{\substack{j=1 \ j \neq i}}^{n} \left[\rho_{B} \left(\mathbf{w} \right) \right]_{j} \left[\mathbf{e}_{j} \right]_{i}$$
 Property CC [205]
$$= \left[\rho_{B} \left(\mathbf{w} \right) \right]_{i} \left(1 \right) + \sum_{\substack{j=1 \ j \neq i}}^{n} \left[\rho_{B} \left(\mathbf{w} \right) \right]_{j} \left(0 \right)$$
 Definition SUV [409]
$$= \left[\rho_{B} \left(\mathbf{w} \right) \right]_{i}$$

As column vectors, Definition CVE [199] implies that $T(\mathbf{w}) = \rho_B(\mathbf{w})$. Since \mathbf{w} was an arbitrary element of V, as functions $T = \rho_B$. Now, since T is known to be a linear transformation, it must follow that ρ_B is also a linear transformation.

The proof of Theorem VRLT [1262] provides an alternate definition of vector representation relative to a basis B that we could state as a corollary (Technique LC [1634]): ρ_B is the unique linear transformation that takes B to the standard unit basis.

Example VRC4 Vector representation in \mathbb{C}^4

Consider the vector $\mathbf{y} \in \mathbb{C}^4$

$$\mathbf{y} = \begin{bmatrix} 6\\14\\6\\7\end{bmatrix}$$

We will find several vector representations of \mathbf{y} in this example. Notice that \mathbf{y} never changes, but the *representations* of \mathbf{y} do change.

One basis for \mathbb{C}^4 is

$$B = \{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3, \, \mathbf{u}_4\} = \left\{ \begin{bmatrix} -2\\1\\2\\-3 \end{bmatrix}, \begin{bmatrix} 3\\-6\\2\\-4 \end{bmatrix}, \begin{bmatrix} 1\\2\\0\\5 \end{bmatrix}, \begin{bmatrix} 4\\3\\1\\6 \end{bmatrix} \right\}$$

as can be seen by making these vectors the columns of a matrix, checking that the matrix is nonsingular and applying Theorem CNMB [789]. To find $\rho_B(\mathbf{y})$, we need to find scalars, a_1, a_2, a_3, a_4 such that

$$\mathbf{y} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + a_4 \mathbf{u}_4$$

By Theorem SLSLC [229] the desired scalars are a solution to the linear system of equations with a coefficient matrix whose columns are the vectors in B and with a vector of constants **y**. With a nonsingular coefficient matrix, the solution is unique, but this is no surprise as this is the content of Theorem VRRB [755]. This unique solution is

$$a_1 = 2$$
 $a_2 = -1$ $a_3 = -3$ $a_4 = 4$

Then by Definition VR [1261], we have

$$\rho_B\left(\mathbf{y}\right) = \begin{bmatrix} 2\\ -1\\ -3\\ 4 \end{bmatrix}$$

Suppose now that we construct a representation of \mathbf{y} relative to another basis of \mathbb{C}^4 ,

$$C = \left\{ \begin{bmatrix} -15\\9\\-4\\-2 \end{bmatrix}, \begin{bmatrix} 16\\-14\\5\\2 \end{bmatrix}, \begin{bmatrix} -26\\14\\-6\\-3 \end{bmatrix}, \begin{bmatrix} 14\\-13\\4\\6 \end{bmatrix} \right\}$$

As with B, it is easy to check that C is a basis. Writing **y** as a linear combination of the vectors in C leads to solving a system of four equations in the four unknown scalars with a nonsingular coefficient matrix. The unique solution can be expressed as

$$\mathbf{y} = \begin{bmatrix} 6\\14\\6\\7 \end{bmatrix} = (-28) \begin{bmatrix} -15\\9\\-4\\-2 \end{bmatrix} + (-8) \begin{bmatrix} 16\\-14\\5\\2 \end{bmatrix} + 11 \begin{bmatrix} -26\\14\\-6\\-3 \end{bmatrix} + 0 \begin{bmatrix} 14\\-13\\4\\6 \end{bmatrix}$$

so that Definition VR [1261] gives

$$\rho_C\left(\mathbf{y}\right) = \begin{bmatrix} -28\\ -8\\ 11\\ 0 \end{bmatrix}$$

We often perform representations relative to standard bases, but for vectors in \mathbb{C}^m it's a little silly. Let's find the vector representation of **y** relative to the standard basis (Theorem SUVB [779]),

$$D = \{\mathbf{e}_1, \, \mathbf{e}_2, \, \mathbf{e}_3, \, \mathbf{e}_4\}$$

Then, without any computation, we can check that

$$\mathbf{y} = \begin{bmatrix} 6\\14\\6\\7 \end{bmatrix} = 6\mathbf{e}_1 + 14\mathbf{e}_2 + 6\mathbf{e}_3 + 7\mathbf{e}_4$$

so by Definition VR [1261],

$$\rho_D\left(\mathbf{y}\right) = \begin{bmatrix} 6\\14\\6\\7\end{bmatrix}$$

which is not very exciting. Notice however that the *order* in which we place the vectors in the basis is critical to the representation. Let's keep the standard unit vectors as our basis, but rearrange the order we place them in the basis. So a fourth basis is

$$E = \{\mathbf{e}_3, \, \mathbf{e}_4, \, \mathbf{e}_2, \, \mathbf{e}_1\}$$

Then,

$$\mathbf{y} = \begin{bmatrix} 6\\14\\6\\7 \end{bmatrix} = 6\mathbf{e}_3 + 7\mathbf{e}_4 + 14\mathbf{e}_2 + 6\mathbf{e}_1$$

so by Definition VR [1261],

$$\rho_E\left(\mathbf{y}\right) = \begin{bmatrix} 6\\7\\14\\6 \end{bmatrix}$$

So for every possible basis of \mathbb{C}^4 we could construct a different representation of **y**.

Vector representations are most interesting for vector spaces that are not \mathbb{C}^m .

Example VRP2

Vector representations in P_2

Consider the vector $\mathbf{u} = 15 + 10x - 6x^2 \in P_2$ from the vector space of polynomials with

degree at most 2 (Example VSP [665]). A nice basis for P_2 is

$$B = \left\{1, \, x, \, x^2\right\}$$

so that

$$\mathbf{u} = 15 + 10x - 6x^2 = 15(1) + 10(x) + (-6)(x^2)$$

so by Definition VR [1261]

$$\rho_B\left(\mathbf{u}\right) = \begin{bmatrix} 15\\10\\-6 \end{bmatrix}$$

Another nice basis for P_2 is

$$C = \left\{1, 1+x, 1+x+x^2\right\}$$

so that now it takes a bit of computation to determine the scalars for the representation. We want a_1 , a_2 , a_3 so that

$$15 + 10x - 6x^{2} = a_{1}(1) + a_{2}(1+x) + a_{3}(1+x+x^{2})$$

Performing the operations in P_2 on the right-hand side, and equating coefficients, gives the three equations in the three unknown scalars,

 $15 = a_1 + a_2 + a_3$ $10 = a_2 + a_3$ $-6 = a_3$ The coefficient matrix of this system is nonsingular, leading to a unique solution (no surprise there, see Theorem VRRB [755]),

$$a_1 = 5$$
 $a_2 = 16$ $a_3 = -6$

so by Definition VR [1261]

$$\rho_C\left(\mathbf{u}\right) = \begin{bmatrix} 5\\16\\-6 \end{bmatrix}$$

While we often form vector representations relative to "nice" bases, nothing prevents us from forming representations relative to "nasty" bases. For example, the set

$$D = \left\{-2 - x + 3x^2, \ 1 - 2x^2, \ 5 + 4x + x^2\right\}$$

can be verified as a basis of P_2 by checking linear independence with Definition LI [735] and then arguing that 3 vectors from P_2 , a vector space of dimension 3 (Theorem DP [824]), must also be a spanning set (Theorem G [852]). Now we desire scalars a_1 , a_2 , a_3 so that

$$15 + 10x - 6x^{2} = a_{1}(-2 - x + 3x^{2}) + a_{2}(1 - 2x^{2}) + a_{3}(5 + 4x + x^{2})$$

Performing the operations in P_2 on the right-hand side, and equating coefficients, gives the three equations in the three unknown scalars,

$$15 = -2a_1 + a_2 + 5a_3$$

 $10 = -a_1 + 4a_3$ -6 = 3a_1 - 2a_2 + a_3

The coefficient matrix of this system is nonsingular, leading to a unique solution (no surprise there, see Theorem VRRB [755]),

 $a_1 = -2$ $a_2 = 1$ $a_3 = 2$

so by Definition VR [1261]

$$\rho_D\left(\mathbf{u}\right) = \begin{bmatrix} -2\\1\\2 \end{bmatrix}$$

Theorem VRI Vector Representation is Injective

The function ρ_B (Definition VR [1261]) is an injective linear transformation.

Proof We will appeal to Theorem KILT [1144]. Suppose U is a vector space of dimension n, so vector representation is of the form $\rho_B \colon U \to \mathbb{C}^n$. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n\}$ be the basis of U used in the definition of ρ_B . Suppose $\mathbf{u} \in \mathcal{K}(\rho_B)$. We write \mathbf{u} as a linear combination of the vectors in the basis B where the scalars are the components of the vector representation, $\rho_B(\mathbf{u})$.

$$\mathbf{u} = \left[\rho_B\left(\mathbf{u}\right)\right]_1 \mathbf{u}_1 + \left[\rho_B\left(\mathbf{u}\right)\right]_2 \mathbf{u}_2 + \left[\rho_B\left(\mathbf{u}\right)\right]_3 \mathbf{u}_3 + \dots + \left[\rho_B\left(\mathbf{u}\right)\right]_n \mathbf{u}_n \quad \text{Definition VR} \ [1261]$$

 \boxtimes

$$= [\mathbf{0}]_{1} \mathbf{u}_{1} + [\mathbf{0}]_{2} \mathbf{u}_{2} + [\mathbf{0}]_{3} \mathbf{u}_{3} + \dots + [\mathbf{0}]_{n} \mathbf{u}_{n}$$

$$= 0 \mathbf{u}_{1} + 0 \mathbf{u}_{2} + 0 \mathbf{u}_{3} + \dots + 0 \mathbf{u}_{n}$$

$$= \mathbf{0} + \mathbf{0} + \mathbf{0} + \dots + \mathbf{0}$$

$$= \mathbf{0}$$
Definition KLT [1138]
Definition ZCV [53]
Theorem ZSSM [676]
Property Z [660]

Thus an arbitrary vector, **u**, from the kernel $\mathcal{K}(\rho_B)$, must equal the zero vector of U. So $\mathcal{K}(\rho_B) = \{\mathbf{0}\}$ and by Theorem KILT [1144], ρ_B is injective.

Theorem VRS Vector Representation is Surjective

The function ρ_B (Definition VR [1261]) is a surjective linear transformation.

Proof We will appeal to Theorem RSLT [1185]. Suppose U is a vector space of dimension n, so vector representation is of the form $\rho_B \colon U \to \mathbb{C}^n$. Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n\}$ be the basis of U used in the definition of ρ_B . Suppose $\mathbf{v} \in \mathbb{C}^n$. Define the vector \mathbf{u} by

$$\mathbf{u} = [\mathbf{v}]_1 \, \mathbf{u}_1 + [\mathbf{v}]_2 \, \mathbf{u}_2 + [\mathbf{v}]_3 \, \mathbf{u}_3 + \dots + [\mathbf{v}]_n \, \mathbf{u}_n$$

Then for $1 \leq i \leq n$

$$[\rho_B (\mathbf{u})]_i = [\rho_B ([\mathbf{v}]_1 \mathbf{u}_1 + [\mathbf{v}]_2 \mathbf{u}_2 + [\mathbf{v}]_3 \mathbf{u}_3 + \dots + [\mathbf{v}]_n \mathbf{u}_n)]_i$$

= $[\mathbf{v}]_i$ Definition VR [1261]

so the entries of vectors $\rho_B(\mathbf{u})$ and \mathbf{v} are equal and Definition CVE [199] yields the vector equality $\rho_B(\mathbf{u}) = \mathbf{v}$. This demonstrates that $\mathbf{v} \in \mathcal{R}(\rho_B)$, so $\mathbb{C}^n \subseteq \mathcal{R}(\rho_B)$. Since $\mathcal{R}(\rho_B) \subseteq \mathbb{C}^n$

 \square

by Definition RLT [1178], we have $\mathcal{R}(\rho_B) = \mathbb{C}^n$ and Theorem RSLT [1185] says ρ_B is surjective.

We will have many occasions later to employ the inverse of vector representation, so we will record the fact that vector representation is an invertible linear transformation.

Theorem VRILT Vector Representation is an Invertible I

Vector Representation is an Invertible Linear Transformation

The function ρ_B (Definition VR [1261]) is an invertible linear transformation.

Proof The function ρ_B (Definition VR [1261]) is a linear transformation (Theorem VRLT [1262]) that is injective (Theorem VRI [1270]) and surjective (Theorem VRS [1271]) with domain V and codomain \mathbb{C}^n . By Theorem ILTIS [1221] we then know that ρ_B is an invertible linear transformation.

Informally, we will refer to the application of ρ_B as **coordinatizing** a vector, while the application of ρ_B^{-1} will be referred to as **un-coordinatizing** a vector.

Subsection CVS Characterization of Vector Spaces

Limiting our attention to vector spaces with finite dimension, we now describe every possible vector space. All of them. Really.

Theorem CFDVS Characterization of Finite Dimensional Vector Spaces

Suppose that V is a vector space with dimension n. Then V is isomorphic to \mathbb{C}^n .

Proof Since V has dimension n we can find a basis of V of size n (Definition D [815]) which we will call B. The linear transformation ρ_B is an invertible linear transformation from V to \mathbb{C}^n , so by Definition IVS [1230], we have that V and \mathbb{C}^n are isomorphic.

Theorem CFDVS [1272] is the first of several surprises in this chapter, though it might be a bit demoralizing too. It says that there really are not all that many different (finite dimensional) vector spaces, and none are really any more complicated than \mathbb{C}^n . Hmmm. The following examples should make this point.

Example TIVS Two isomorphic vector spaces

The vector space of polynomials with degree 8 or less, P_8 , has dimension 9 (Theorem DP [824]). By Theorem CFDVS [1272], P_8 is isomorphic to \mathbb{C}^9 .

Example CVSR

Crazy vector space revealed

The crazy vector space, C of Example CVS [670], has dimension 2 by Example DC [827]. By Theorem CFDVS [1272], C is isomorphic to \mathbb{C}^2 . Hmmmm. Not really so crazy after all?

Example ASC

A subspace characterized

In Example DSP4 [827] we determined that a certain subspace W of P_4 has dimension 4. By Theorem CFDVS [1272], W is isomorphic to \mathbb{C}^4 .

Theorem IFDVS

Isomorphism of Finite Dimensional Vector Spaces

Suppose U and V are both finite-dimensional vector spaces. Then U and V are isomorphic if and only if $\dim(U) = \dim(V)$.

Proof (\Rightarrow) This is just the statement proved in Theorem IVSED [1232].

(\Leftarrow) This is the advertised converse of Theorem IVSED [1232]. We will assume U and V have equal dimension and discover that they are isomorphic vector spaces. Let n be the common dimension of U and V. Then by Theorem CFDVS [1272] there are isomorphisms $T: U \to \mathbb{C}^n$ and $S: V \to \mathbb{C}^n$.

T is therefore an invertible linear transformation by Definition IVS [1230]. Similarly, S is an invertible linear transformation, and so S^{-1} is an invertible linear transformation (Theorem IILT [1220]). The composition of invertible linear transformations is again invertible (Theorem CIVLT [1227]) so the composition of S^{-1} with T is invertible. Then $(S^{-1} \circ T) : U \to V$ is an invertible linear transformation from U to V and Definition IVS [1230] says U and V are isomorphic.

Example MIVS Multiple isomorphic vector spaces

 \mathbb{C}^{10} , P_9 , $M_{2,5}$ and $M_{5,2}$ are all vector spaces and each has dimension 10. By Theorem IFDVS

[1274] each is isomorphic to any other.

The subspace of $M_{4,4}$ that contains all the symmetric matrices (Definition SYM [434]) has dimension 10, so this subspace is also isomorphic to each of the four vector spaces above.

Subsection CP Coordinatization Principle

With ρ_B available as an invertible linear transformation, we can translate between vectors in a vector space U of dimension m and \mathbb{C}^m . Furthermore, as a linear transformation, ρ_B respects the addition and scalar multiplication in U, while ρ_B^{-1} respects the addition and scalar multiplication in \mathbb{C}^m . Since our definitions of linear independence, spans, bases and dimension are all built up from linear combinations, we will finally be able to translate fundamental properties between abstract vector spaces (U) and concrete vector spaces (\mathbb{C}^m) .

Theorem CLI

Coordinatization and Linear Independence

Suppose that U is a vector space with a basis B of size n. Then $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k\}$ is a linearly independent subset of U if and only if $R = \{\rho_B(\mathbf{u}_1), \rho_B(\mathbf{u}_2), \rho_B(\mathbf{u}_3), \ldots, \rho_B(\mathbf{u}_k)\}$ is a linearly independent subset of \mathbb{C}^n .

Proof The linear transformation ρ_B is an isomorphism between U and \mathbb{C}^n (Theorem VRILT

[1272]). As an invertible linear transformation, ρ_B is an injective linear transformation (Theorem ILTIS [1221]), and ρ_B^{-1} is also an injective linear transformation (Theorem IILT [1220], Theorem ILTIS [1221]).

 (\Rightarrow) Since ρ_B is an injective linear transformation and S is linearly independent, Theorem ILTLI [1148] says that R is linearly independent.

(⇐) If we apply ρ_B^{-1} to each element of R, we will create the set S. Since we are assuming R is linearly independent and ρ_B^{-1} is injective, Theorem ILTLI [1148] says that S is linearly independent.

Theorem CSS

Coordinatization and Spanning Sets

Suppose that U is a vector space with a basis B of size n. Then $\mathbf{u} \in \langle \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_k\} \rangle$ if and only if $\rho_B(\mathbf{u}) \in \langle \{\rho_B(\mathbf{u}_1), \rho_B(\mathbf{u}_2), \rho_B(\mathbf{u}_3), \dots, \rho_B(\mathbf{u}_k)\} \rangle$.

Proof (\Rightarrow) Suppose $\mathbf{u} \in \langle \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k\} \rangle$. Then there are scalars, $a_1, a_2, a_3, \ldots, a_k$, such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_k\mathbf{u}_k$$

Then,

$$\rho_B \left(\mathbf{u} \right) = \rho_B \left(a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_k \mathbf{u}_k \right)$$

= $a_1 \rho_B \left(\mathbf{u}_1 \right) + a_2 \rho_B \left(\mathbf{u}_2 \right) + a_3 \rho_B \left(\mathbf{u}_3 \right) + \dots + a_k \rho_B \left(\mathbf{u}_k \right)$ Theorem LTLC [1093]

which says that $\rho_B(\mathbf{u}) \in \langle \{\rho_B(\mathbf{u}_1), \rho_B(\mathbf{u}_2), \rho_B(\mathbf{u}_3), \dots, \rho_B(\mathbf{u}_k) \} \rangle$.

(\Leftarrow) Suppose that $\rho_B(\mathbf{u}) \in \langle \{\rho_B(\mathbf{u}_1), \rho_B(\mathbf{u}_2), \rho_B(\mathbf{u}_3), \dots, \rho_B(\mathbf{u}_k) \} \rangle$. Then there are scalars $b_1, b_2, b_3, \dots, b_k$ such that

$$\rho_B\left(\mathbf{u}\right) = b_1 \rho_B\left(\mathbf{u}_1\right) + b_2 \rho_B\left(\mathbf{u}_2\right) + b_3 \rho_B\left(\mathbf{u}_3\right) + \dots + b_k \rho_B\left(\mathbf{u}_k\right)$$

Recall that ρ_B is invertible (Theorem VRILT [1272]), so

$$\begin{aligned} \mathbf{u} &= I_U \left(\mathbf{u} \right) & \text{Definition IDLT [1213]} \\ &= \left(\rho_B^{-1} \circ \rho_B \right) \left(\mathbf{u} \right) & \text{Definition IVLT [1214]} \\ &= \rho_B^{-1} \left(\rho_B \left(\mathbf{u} \right) \right) & \text{Definition IVLT [1214]} \\ &= \rho_B^{-1} \left(b_1 \rho_B \left(\mathbf{u}_1 \right) + b_2 \rho_B \left(\mathbf{u}_2 \right) + b_3 \rho_B \left(\mathbf{u}_3 \right) + \dots + b_k \rho_B \left(\mathbf{u}_k \right) \right) \\ &= b_1 \rho_B^{-1} \left(\rho_B \left(\mathbf{u}_1 \right) \right) + b_2 \rho_B^{-1} \left(\rho_B \left(\mathbf{u}_2 \right) \right) + b_3 \rho_B^{-1} \left(\rho_B \left(\mathbf{u}_3 \right) \right) \\ &+ \dots + b_k \rho_B^{-1} \left(\rho_B \left(\mathbf{u}_k \right) \right) & \text{Theorem LTLC [1093]} \\ &= b_1 I_U \left(\mathbf{u}_1 \right) + b_2 I_U \left(\mathbf{u}_2 \right) + b_3 I_U \left(\mathbf{u}_3 \right) + \dots + b_k I_U \left(\mathbf{u}_k \right) & \text{Definition IVLT [1214]} \\ &= b_1 \mathbf{u}_1 + b_2 \mathbf{u}_2 + b_3 \mathbf{u}_3 + \dots + b_k \mathbf{u}_k & \text{Definition IDLT [1213]} \end{aligned}$$

which says that $\mathbf{u} \in \langle \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_k \} \rangle$.

Here's a fairly simple example that illustrates a very, very important idea.

Example CP2 Coordinatizing in P_2 In Example VRP2 [1267] we needed to know that

$$D = \left\{-2 - x + 3x^2, \ 1 - 2x^2, \ 5 + 4x + x^2\right\}$$

is a basis for P_2 . With Theorem CLI [1275] and Theorem CSS [1276] this task is much easier. First, choose a known basis for P_2 , a basis that forms vector representations easily. We will choose

$$B = \left\{1, \, x, \, x^2\right\}$$

Now, form the subset of \mathbb{C}^3 that is the result of applying ρ_B to each element of D,

$$F = \left\{ \rho_B \left(-2 - x + 3x^2 \right), \, \rho_B \left(1 - 2x^2 \right), \, \rho_B \left(5 + 4x + x^2 \right) \right\} = \left\{ \begin{bmatrix} -2 \\ -1 \\ 3 \end{bmatrix}, \, \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}, \, \begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} \right\}$$

and ask if F is a linearly independent spanning set for \mathbb{C}^3 . This is easily seen to be the case by forming a matrix A whose columns are the vectors of F, row-reducing A to the identity matrix I_3 , and then using the nonsingularity of A to assert that F is a basis for \mathbb{C}^3 (Theorem CNMB [789]). Now, since F is a basis for \mathbb{C}^3 , Theorem CLI [1275] and Theorem CSS [1276] tell us that D is also a basis for P_2 .

Example CP2 [1277] illustrates the broad notion that computations in abstract vector spaces can be reduced to computations in \mathbb{C}^m . You may have noticed this phenomenon as you worked through examples in Chapter VS [658] or Chapter LT [1071] employing vector spaces of matrices or polynomials. These computations seemed to invariably result in systems of equations or the like from Chapter SLE [2], Chapter V [197] and Chapter M [425]. It is vector representation, ρ_B , that allows us to make this connection formal and precise.

Knowing that vector representation allows us to translate questions about linear combinations, linear independence and spans from general vector spaces to \mathbb{C}^m allows us to prove a great many theorems about how to translate other properties. Rather than prove these theorems, each of the same style as the other, we will offer some general guidance about how to best employ Theorem VRLT [1262], Theorem CLI [1275] and Theorem CSS [1276]. This comes in the form of a "principle": a basic truth, but most definitely not a theorem (hence, no proof).

The Coordinatization Principle Suppose that U is a vector space with a basis B of size n. Then any question about U, or its elements, which ultimately depends on the vector addition or scalar multiplication in U, or depends on linear independence or spanning, may be translated into the same question in \mathbb{C}^n by application of the linear transformation ρ_B to the relevant vectors. Once the question is answered in \mathbb{C}^n , the answer may be translated back to U through application of the inverse linear transformation ρ_B^{-1} (if necessary).

Example CM32 Coordinatization in M_{32}

This is a simple example of the Coordinatization Principle [1279], depending only on the fact that coordinatizing is an invertible linear transformation (Theorem VRILT [1272]). Suppose we have a linear combination to perform in M_{32} , the vector space of 3×2 matrices, but we are adverse to doing the operations of M_{32} (Definition MA [427], Definition MSM [428]). More specifically, suppose we are faced with the computation

$$6\begin{bmatrix} 3 & 7\\ -2 & 4\\ 0 & -3 \end{bmatrix} + 2\begin{bmatrix} -1 & 3\\ 4 & 8\\ -2 & 5 \end{bmatrix}$$

We choose a nice basis for M_{32} (or a nasty basis if we are so inclined),

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

and apply ρ_B to each vector in the linear combination. This gives us a new computation, now in the vector space \mathbb{C}^6 ,

$$6\begin{bmatrix} 3\\ -2\\ 0\\ 7\\ 4\\ -3\end{bmatrix} + 2\begin{bmatrix} -1\\ 4\\ -2\\ 3\\ 8\\ 5\end{bmatrix}$$

which we can compute with the operations of \mathbb{C}^6 (Definition CVA [201], Definition CVSM [203]), to arrive at

$$\begin{bmatrix} 16 \\ -4 \\ -4 \\ 48 \\ 40 \\ -8 \end{bmatrix}$$

We are after the result of a computation in M_{32} , so we now can apply ρ_B^{-1} to obtain a 3 \times 2

matrix,

$$16\begin{bmatrix}1 & 0\\0 & 0\\0 & 0\end{bmatrix} + (-4)\begin{bmatrix}0 & 0\\1 & 0\\0 & 0\end{bmatrix} + (-4)\begin{bmatrix}0 & 0\\0 & 0\\1 & 0\end{bmatrix} + 48\begin{bmatrix}0 & 1\\0 & 0\\0 & 0\end{bmatrix} + 40\begin{bmatrix}0 & 0\\0 & 1\\0 & 0\end{bmatrix} + (-8)\begin{bmatrix}0 & 0\\0 & 0\\0 & 1\end{bmatrix} = \begin{bmatrix}16 & 48\\-4 & 40\\-4 & -8\end{bmatrix}$$

which is exactly the matrix we would have computed had we just performed the matrix operations in the first place. So this was not meant to be an *easier* way to compute a linear combination of two matrices, just a *different* way. \boxtimes

Subsection READ Reading Questions

- 1. The vector space of 3×5 matrices, $M_{3,5}$ is isomorphic to what fundamental vector space?
- 2. A basis for \mathbb{C}^3 is

$$B = \left\{ \begin{bmatrix} 1\\2\\-1 \end{bmatrix}, \begin{bmatrix} 3\\-1\\2 \end{bmatrix}, \begin{bmatrix} 1\\1\\1 \end{bmatrix} \right\}$$

Compute
$$\rho_B \left(\begin{bmatrix} 5\\8\\-1 \end{bmatrix} \right)$$
.

3. What is the first "surprise," and why is it surprising?

Subsection EXC Exercises

C10 In the vector space \mathbb{C}^3 , compute the vector representation $\rho_B(\mathbf{v})$ for the basis *B* and vector \mathbf{v} below.

$$B = \left\{ \begin{bmatrix} 2\\-2\\2 \end{bmatrix}, \begin{bmatrix} 1\\3\\1 \end{bmatrix}, \begin{bmatrix} 3\\5\\2 \end{bmatrix} \right\} \qquad \mathbf{v} = \begin{bmatrix} 11\\5\\8 \end{bmatrix}$$

Contributed by Robert Beezer Solution [1284]

C20 Rework Example CM32 [1279] replacing the basis B by the basis

$$C = \left\{ \begin{bmatrix} -14 & -9\\ 10 & 10\\ -6 & -2 \end{bmatrix}, \begin{bmatrix} -7 & -4\\ 5 & 5\\ -3 & -1 \end{bmatrix}, \begin{bmatrix} -3 & -1\\ 0 & -2\\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -7 & -4\\ 3 & 2\\ -1 & 0 \end{bmatrix}, \begin{bmatrix} 4 & 2\\ -3 & -3\\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0\\ -1 & -2\\ 1 & 1 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [1285]

M10 Prove that the set S below is a basis for the vector space of 2×2 matrices, M_{22} . Do this choosing a natural basis for M_{22} and coordinatizing the elements of S with respect to this

basis. Examine the resulting set of column vectors from \mathbb{C}^4 and apply the Coordinatization Principle [1279].

$$S = \left\{ \begin{bmatrix} 33 & 99\\ 78 & -9 \end{bmatrix}, \begin{bmatrix} -16 & -47\\ -36 & 2 \end{bmatrix}, \begin{bmatrix} 10 & 27\\ 17 & 3 \end{bmatrix}, \begin{bmatrix} -2 & -7\\ -6 & 4 \end{bmatrix} \right\}$$

Contributed by Andy Zimmer

Subsection SOL Solutions

C10 Contributed by Robert Beezer Statement [1282] We need to express the vector \mathbf{v} as a linear combination of the vectors in B. Theorem VRRB [755] tells us we will be able to do this, and do it uniquely. The vector equation

$$a_1 \begin{bmatrix} 2\\-2\\2 \end{bmatrix} + a_2 \begin{bmatrix} 1\\3\\1 \end{bmatrix} + a_3 \begin{bmatrix} 3\\5\\2 \end{bmatrix} = \begin{bmatrix} 11\\5\\8 \end{bmatrix}$$

becomes (via Theorem SLSLC [229]) a system of linear equations with augmented matrix,

$\begin{bmatrix} 2 \end{bmatrix}$	1	3	11]
-2	3	5	5
2	1	2	8

This system has the unique solution $a_1 = 2$, $a_2 = -2$, $a_3 = 3$. So by Definition VR [1261],

$$\rho_B\left(\mathbf{v}\right) = \rho_B\left(\begin{bmatrix}11\\5\\8\end{bmatrix}\right) = \rho_B\left(2\begin{bmatrix}2\\-2\\2\end{bmatrix} + (-2)\begin{bmatrix}1\\3\\1\end{bmatrix} + 3\begin{bmatrix}3\\5\\2\end{bmatrix}\right) = \begin{bmatrix}2\\-2\\3\end{bmatrix}$$

C20 Contributed by Robert Beezer Statement [1282]

The following computations replicate the computations given in Example CM32 [1279], only using the basis C.

$$\rho_{C}\left(\begin{bmatrix}3 & 7\\-2 & 4\\0 & -3\end{bmatrix}\right) = \begin{bmatrix}-9\\12\\-6\\7\\-2\\-1\end{bmatrix} \qquad \rho_{C}\left(\begin{bmatrix}-1 & 3\\4 & 8\\-2 & 5\end{bmatrix}\right) = \begin{bmatrix}-11\\34\\-4\\-1\\16\\5\end{bmatrix}$$
$$6\begin{bmatrix}-9\\12\\-6\\7\\-2\\-1\end{bmatrix} + 2\begin{bmatrix}-11\\34\\-4\\-4\\-1\\16\\5\end{bmatrix} = \begin{bmatrix}-76\\140\\-44\\40\\20\\4\end{bmatrix} \qquad \rho_{C}^{-1}\left(\begin{bmatrix}-76\\140\\-44\\40\\20\\4\end{bmatrix}\right) = \begin{bmatrix}16 & 48\\-4 & 30\\-4 & -8\end{bmatrix}$$

Section MR Matrix Representations

We have seen that linear transformations whose domain and codomain are vector spaces of columns vectors have a close relationship with matrices (Theorem MBLT [1086], Theorem MLTCV [1089]). In this section, we will extend the relationship between matrices and linear transformations to the setting of linear transformations between abstract vector spaces.

Definition MR

Matrix Representation

Suppose that $T: U \to V$ is a linear transformation, $B = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n}$ is a basis for U of size n, and C is a basis for V of size m. Then the **matrix representation** of T relative to B and C is the $m \times n$ matrix,

$$M_{B,C}^{T} = \left[\rho_{C}\left(T\left(\mathbf{u}_{1}\right)\right) \middle| \rho_{C}\left(T\left(\mathbf{u}_{2}\right)\right) \middle| \rho_{C}\left(T\left(\mathbf{u}_{3}\right)\right) \middle| \dots \left|\rho_{C}\left(T\left(\mathbf{u}_{n}\right)\right)\right]$$

(This definition contains Notation MR.)

Example OLTTR One linear transformation, three representations \wedge

Consider the linear transformation

ł

$$S: P_3 \to M_{22}, \quad S\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix} 3a + 7b - 2c - 5d & 8a + 14b - 2c - 11d \\ -4a - 8b + 2c + 6d & 12a + 22b - 4c - 17d \end{bmatrix}$$

First, we build a representation relative to the bases,

$$B = \left\{ 1 + 2x + x^2 - x^3, \ 1 + 3x + x^2 + x^3, \ -1 - 2x + 2x^3, \ 2 + 3x + 2x^2 - 5x^3 \right\}$$
$$C = \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix}, \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} -1 & -4 \\ -2 & -4 \end{bmatrix} \right\}$$

We evaluate S with each element of the basis for the domain, B, and coordinatize the result relative to the vectors in the basis for the codomain, C. Notice here how we take elements of vector spaces and decompose them into linear combinations of basis elements as the key step in constructing coordinatizations of vectors. There is a system of equations involved almost every time, but we will omit these details since this should be a routine exercise at this stage.

$$\rho_C \left(S \left(1 + 2x + x^2 - x^3 \right) \right) = \rho_C \left(\begin{bmatrix} 20 & 45 \\ -24 & 69 \end{bmatrix} \right)$$
$$= \rho_C \left((-90) \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + 37 \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} + (-40) \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix} + 4 \begin{bmatrix} -1 & -4 \\ -2 & -4 \end{bmatrix} \right) = \begin{bmatrix} -90 \\ 37 \\ -40 \\ 4 \end{bmatrix}$$

$$\begin{aligned} \rho_C \left(S \left(1 + 3x + x^2 + x^3 \right) \right) &= \rho_C \left(\begin{bmatrix} 17 & 37 \\ -20 & 57 \end{bmatrix} \right) \\ &= \rho_C \left((-72) \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + 29 \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} + (-34) \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix} + 3 \begin{bmatrix} -1 & -4 \\ -2 & -4 \end{bmatrix} \right) = \begin{bmatrix} -72 \\ 29 \\ -34 \\ 3 \end{bmatrix} \\ \rho_C \left(S \left(-1 - 2x + 2x^3 \right) \right) &= \rho_C \left(\begin{bmatrix} -27 & -58 \\ 32 & -90 \end{bmatrix} \right) \\ &= \rho_C \left(114 \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + (-46) \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} + 54 \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix} + (-5) \begin{bmatrix} -1 & -4 \\ -2 & -4 \end{bmatrix} \right) = \begin{bmatrix} 114 \\ -46 \\ 54 \\ -5 \end{bmatrix} \\ \rho_C \left(S \left(2 + 3x + 2x^2 - 5x^3 \right) \right) &= \rho_C \left(\begin{bmatrix} 48 & 109 \\ -58 & 167 \end{bmatrix} \right) \\ &= \rho_C \left((-220) \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + 91 \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix} + -96 \begin{bmatrix} -1 & -1 \\ 0 & -2 \end{bmatrix} + 10 \begin{bmatrix} -1 & -4 \\ -2 & -4 \end{bmatrix} \right) = \begin{bmatrix} -220 \\ 91 \\ -96 \\ 10 \end{bmatrix} \end{aligned}$$

Thus, employing Definition MR [1286]

$$M_{B,C}^{S} = \begin{bmatrix} -90 & -72 & 114 & -220 \\ 37 & 29 & -46 & 91 \\ -40 & -34 & 54 & -96 \\ 4 & 3 & -5 & 10 \end{bmatrix}$$

Often we use "nice" bases to build matrix representations and the work involved is much easier. Suppose we take bases

$$D = \{1, x, x^2, x^3\} \qquad E = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

The evaluation of S at the elements of D is easy and coordinatization relative to E can be done on sight,

$$\rho_E(S(1)) = \rho_E\left(\begin{bmatrix}3 & 8\\-4 & 12\end{bmatrix}\right) \\ = \rho_E\left(3\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 8\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + (-4)\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 12\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}3\\8\\-4\\12\end{bmatrix} \\ \rho_E(S(x)) = \rho_E\left(\begin{bmatrix}7 & 14\\-8 & 22\end{bmatrix}\right)$$

$$= \rho_E \left(7 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 14 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + (-8) \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 22 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 7 \\ 14 \\ -8 \\ 22 \end{bmatrix}$$

$$\rho_E \left(S \left(x^2 \right) \right) = \rho_E \left(\begin{bmatrix} -2 & -2 \\ 2 & -4 \end{bmatrix} \right)$$

$$= \rho_E \left((-2) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (-2) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 2 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + (-4) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} -2 \\ -2 \\ 2 \\ -4 \end{bmatrix}$$

$$\rho_E \left(S \left(x^3 \right) \right) = \rho_E \left(\begin{bmatrix} -5 & -11 \\ 6 & -17 \end{bmatrix} \right)$$

$$= \rho_E \left((-5) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (-11) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 6 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + (-17) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} -5 \\ -11 \\ 6 \\ -17 \end{bmatrix}$$

So the matrix representation of S relative to D and E is

$$M_{D,E}^{S} = \begin{bmatrix} 3 & 7 & -2 & -5 \\ 8 & 14 & -2 & -11 \\ -4 & -8 & 2 & 6 \\ 12 & 22 & -4 & -17 \end{bmatrix}$$

One more time, but now let's use bases

$$F = \left\{ 1 + x - x^2 + 2x^3, -1 + 2x + 2x^3, 2 + x - 2x^2 + 3x^3, 1 + x + 2x^3 \right\}$$
$$G = \left\{ \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} -1 & 2 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ -2 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \right\}$$

and evaluate S with the elements of F, then coordinatize the results relative to G,

$$\rho_G \left(S \left(1 + x - x^2 + 2x^3 \right) \right) = \rho_G \left(\begin{bmatrix} 2 & 2 \\ -2 & 4 \end{bmatrix} \right) = \rho_G \left(2 \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_G \left(S \left(-1 + 2x + 2x^3 \right) \right) = \rho_G \left(\begin{bmatrix} 1 & -2 \\ 0 & -2 \end{bmatrix} \right) = \rho_G \left((-1) \begin{bmatrix} -1 & 2 \\ 0 & 2 \end{bmatrix} \right) = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

$$\rho_G\left(S\left(2+x-2x^2+3x^3\right)\right) = \rho_G\left(\begin{bmatrix}2&1\\-2&3\end{bmatrix}\right) = \rho_G\left(\begin{bmatrix}2&1\\-2&3\end{bmatrix}\right) = \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$$
$$\rho_G\left(S\left(1+x+2x^3\right)\right) = \rho_G\left(\begin{bmatrix}0&0\\0&0\end{bmatrix}\right) = \rho_G\left(0\begin{bmatrix}1&1\\0&2\end{bmatrix}\right) = \begin{bmatrix}0\\0\\0\\0\end{bmatrix}$$

So we arrive at an especially economical matrix representation,

$$M_{F,G}^{S} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

 \boxtimes

We may choose to use whatever terms we want when we make a definition. Some are arbitrary, while others make sense, but only in light of subsequent theorems. Matrix representation is in the latter category. We begin with a linear transformation and produce a matrix. So what? Here's the theorem that *justifies* the term "matrix representation."

Theorem FTMR

Fundamental Theorem of Matrix Representation

Suppose that $T: U \to V$ is a linear transformation, B is a basis for U, C is a basis for Vand $M_{B,C}^T$ is the matrix representation of T relative to B and C. Then, for any $\mathbf{u} \in U$,

$$\rho_{C}\left(T\left(\mathbf{u}\right)\right) = M_{B,C}^{T}\left(\rho_{B}\left(\mathbf{u}\right)\right)$$

or equivalently

$$T\left(\mathbf{u}\right) = \rho_{C}^{-1}\left(M_{B,C}^{T}\left(\rho_{B}\left(\mathbf{u}\right)\right)\right)$$

Proof Let $B = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n}$ be the basis of U. Since $\mathbf{u} \in U$, there are scalars $a_1, a_2, a_3, \ldots, a_n$ such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \dots + a_n\mathbf{u}_n$$

Then,

$$M_{B,C}^{T}\rho_{B}\left(\mathbf{u}\right)$$

$$= \left[\rho_{C}\left(T\left(\mathbf{u}_{1}\right)\right) | \rho_{C}\left(T\left(\mathbf{u}_{2}\right)\right) | \rho_{C}\left(T\left(\mathbf{u}_{3}\right)\right) | \dots | \rho_{C}\left(T\left(\mathbf{u}_{n}\right)\right) \right] \rho_{B}\left(\mathbf{u}\right) \quad \text{Definition MR [1286]}$$

$$= \left[\rho_{C}\left(T\left(\mathbf{u}_{1}\right)\right) | \rho_{C}\left(T\left(\mathbf{u}_{2}\right)\right) | \rho_{C}\left(T\left(\mathbf{u}_{3}\right)\right) | \dots | \rho_{C}\left(T\left(\mathbf{u}_{n}\right)\right) \right] \begin{bmatrix}a_{1}\\a_{2}\\a_{3}\\\vdots\\a_{n}\end{bmatrix} \quad \text{Definition VR [1261]}$$

$$= a_1 \rho_C \left(T \left(\mathbf{u}_1 \right) \right) + a_2 \rho_C \left(T \left(\mathbf{u}_2 \right) \right) + \dots + a_n \rho_C \left(T \left(\mathbf{u}_n \right) \right)$$

$$= \rho_C \left(a_1 T \left(\mathbf{u}_1 \right) + a_2 T \left(\mathbf{u}_2 \right) + a_3 T \left(\mathbf{u}_3 \right) + \dots + a_n T \left(\mathbf{u}_n \right) \right)$$

$$= \rho_C \left(T \left(a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + a_3 \mathbf{u}_3 + \dots + a_n \mathbf{u}_n \right) \right)$$

$$= \rho_C \left(T \left(\mathbf{u} \right) \right)$$
Definition MVP [461]
Theorem LTLC [1093]
Theorem LTLC [1093]

The alternative conclusion is obtained as

$$T (\mathbf{u}) = I_V (T (\mathbf{u}))$$

= $(\rho_C^{-1} \circ \rho_C) (T (\mathbf{u}))$
= $\rho_C^{-1} (\rho_C (T (\mathbf{u})))$
= $\rho_C^{-1} (M_{B,C}^T (\rho_B (\mathbf{u})))$

Definition IDLT [1213] Definition IVLT [1214] Definition LTC [1111]

This theorem says that we can apply T to **u** and coordinatize the result relative to C in V, or we can first coordinatize **u** relative to B in U, then multiply by the matrix representation. Either way, the result is the same. So the effect of a linear transformation can always be accomplished by a matrix-vector product (Definition MVP [461]). That's important enough to say again. The effect of a linear transformation is a matrix-vector product.

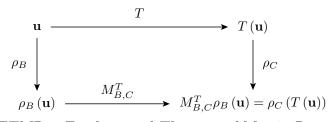


Diagram FTMR. Fundamental Theorem of Matrix Representations

The alternative conclusion of this result might be even more striking. It says that to effect a linear transformation (T) of a vector (\mathbf{u}) , coordinatize the input (with ρ_B), do a matrixvector product (with $M_{B,C}^T$), and un-coordinatize the result (with ρ_C^{-1}). So, absent some bookkeeping about vector representations, a linear transformation *is* a matrix. To adjust the diagram, we "reverse" the arrow on the right, which means inverting the vector representation ρ_C on V. Now we can go directly across the top of the diagram, computing the linear transformation between the abstract vector spaces. Or, we can around the other three sides, using vector representation, a matrix-vector product, followed by un-coordinatization.

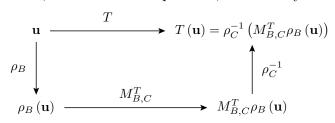


Diagram FTMRA. Fundamental Theorem of Matrix Representations (Alternate)

Here's an example to illustrate how the "action" of a linear transformation can be effected by matrix multiplication.

Example ALTMM

A linear transformation as matrix multiplication

In Example OLTTR [1286] we found three representations of the linear transformation S. In this example, we will compute a single output of S in four different ways. First "normally," then three times over using Theorem FTMR [1293].

Choose $p(x) = 3 - x + 2x^2 - 5x^3$, for no particular reason. Then the straightforward application of S to p(x) yields

$$S(p(x)) = S (3 - x + 2x^2 - 5x^3)$$

= $\begin{bmatrix} 3(3) + 7(-1) - 2(2) - 5(-5) & 8(3) + 14(-1) - 2(2) - 11(-5) \\ -4(3) - 8(-1) + 2(2) + 6(-5) & 12(3) + 22(-1) - 4(2) - 17(-5) \end{bmatrix}$
= $\begin{bmatrix} 23 & 61 \\ -30 & 91 \end{bmatrix}$

Now use the representation of S relative to the bases B and C and Theorem FTMR [1293]. Note that we will employ the following linear combination in moving from the second line to the third,

$$3 - x + 2x^{2} - 5x^{3} = 48(1 + 2x + x^{2} - x^{3}) + (-20)(1 + 3x + x^{2} + x^{3}) + (-1)(-1 - 2x + 2x^{3}) + (-13)(2 + 3x + 2x^{2} - 5x^{3})$$

$$\begin{split} S\left(p(x)\right) &= \rho_{C}^{-1} \left(M_{B,C}^{S} \rho_{B}\left(p(x)\right)\right) \\ &= \rho_{C}^{-1} \left(M_{B,C}^{S} \rho_{B}\left(3 - x + 2x^{2} - 5x^{3}\right)\right) \\ &= \rho_{C}^{-1} \left(M_{B,C}^{S} \begin{bmatrix} 48\\ -20\\ -1\\ -13 \end{bmatrix}\right) \\ &= \rho_{C}^{-1} \left(\begin{bmatrix} -90 & -72 & 114 & -220\\ 37 & 29 & -46 & 91\\ -40 & -34 & 54 & -96\\ 4 & 3 & -5 & 10 \end{bmatrix} \begin{bmatrix} 48\\ -20\\ -1\\ -13 \end{bmatrix}\right) \\ &= \rho_{C}^{-1} \left(\begin{bmatrix} -134\\ 59\\ -46\\ 7 \end{bmatrix}\right) \\ &= \left(-134\right) \begin{bmatrix} 1 & 1\\ 1 & 2 \end{bmatrix} + 59 \begin{bmatrix} 2 & 3\\ 2 & 5 \end{bmatrix} + \left(-46\right) \begin{bmatrix} -1 & -1\\ 0 & -2 \end{bmatrix} + 7 \begin{bmatrix} -1 & -4\\ -2 & -4 \end{bmatrix} \\ &= \begin{bmatrix} 23 & 61\\ -30 & 91 \end{bmatrix} \end{split}$$

Again, but now with "nice" bases like D and E, and the computations are more transparent.

$$S(p(x)) = \rho_E^{-1} \left(M_{D,E}^S \rho_D(p(x)) \right)$$

$$= \rho_E^{-1} \left(M_{D,E}^S \rho_D \left(3 - x + 2x^2 - 5x^3 \right) \right)$$

= $\rho_E^{-1} \left(M_{D,E}^S \rho_D \left(3(1) + (-1)(x) + 2(x^2) + (-5)(x^3) \right) \right)$
= $\rho_E^{-1} \left(M_{D,E}^S \begin{bmatrix} 3 \\ -1 \\ 2 \\ -5 \end{bmatrix} \right)$
= $\rho_E^{-1} \left(\begin{bmatrix} 3 & 7 & -2 & -5 \\ 8 & 14 & -2 & -11 \\ -4 & -8 & 2 & 6 \\ 12 & 22 & -4 & -17 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 2 \\ -5 \end{bmatrix} \right)$
= $\rho_E^{-1} \left(\begin{bmatrix} 23 \\ 61 \\ -30 \\ 91 \end{bmatrix} \right)$
= $23 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 61 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + (-30) \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 91 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
= $\begin{bmatrix} 23 & 61 \\ -30 & 91 \end{bmatrix}$

OK, last time, now with the bases F and G. The coordinatizations will take some work this time, but the matrix-vector product (Definition MVP [461]) (which is the actual action of the linear transformation) will be especially easy, given the diagonal nature of the matrix

representation, $M_{F,G}^S$. Here we go,

$$\begin{split} S\left(p(x)\right) &= \rho_{G}^{-1} \left(M_{F,G}^{S} \rho_{F}\left(p(x)\right)\right) \\ &= \rho_{G}^{-1} \left(M_{F,G}^{S} \rho_{F}\left(3 - x + 2x^{2} - 5x^{3}\right)\right) \\ &= \rho_{G}^{-1} \left(M_{F,G}^{S} \rho_{F}\left(32(1 + x - x^{2} + 2x^{3}) - 7(-1 + 2x + 2x^{3}) - 17(2 + x - 2x^{2} + 3x^{3}) - 2(1 + 2x^{2} + 3x^{3}) - 2(1 +$$

This example is not meant to necessarily illustrate that any one of these four computations is simpler than the others. Instead, it is meant to illustrate the many different ways we can arrive at the same result, with the last three all employing a matrix representation to effect the linear transformation. \boxtimes

We will use Theorem FTMR [1293] frequently in the next few sections. A typical application will feel like the linear transformation T "commutes" with a vector representation, ρ_C , and as it does the transformation morphs into a matrix, $M_{B,C}^T$, while the vector representation changes to a new basis, ρ_B . Or vice-versa.

Subsection NRFO New Representations from Old

In Subsection LT.NLTFO [1105] we built new linear transformations from other linear transformations. Sums, scalar multiples and compositions. These new linear transformations will have matrix representations as well. How do the new matrix representations relate to the old matrix representations? Here are the three theorems.

Theorem MRSLT Matrix Representation of a Sum of Linear Transformations Suppose that $T: U \to V$ and $S: U \to V$ are linear transformations, B is a basis of U and C is a basis of V. Then

$$M_{B,C}^{T+S} = M_{B,C}^{T} + M_{B,C}^{S}$$

Proof Let \mathbf{x} be any vector in \mathbb{C}^n . Define $\mathbf{u} \in U$ by $\mathbf{u} = \rho_B^{-1}(\mathbf{x})$, so $\mathbf{x} = \rho_B(\mathbf{u})$. Then,

$$M_{B,C}^{T+S} \mathbf{x} = M_{B,C}^{T+S} \rho_B (\mathbf{u})$$
Substitution
$$= \rho_C ((T+S) (\mathbf{u}))$$
Theorem FTMR [1293]
$$= \rho_C (T (\mathbf{u}) + S (\mathbf{u}))$$
Definition LTA [1105]
$$= \rho_C (T (\mathbf{u})) + \rho_C (S (\mathbf{u}))$$
Definition LT [1072]
$$= M_{B,C}^T (\rho_B (\mathbf{u})) + M_{B,C}^S (\rho_B (\mathbf{u}))$$
Theorem FTMR [1293]
$$= (M_{B,C}^T + M_{B,C}^S) \rho_B (\mathbf{u})$$
Theorem MMDAA [476]
$$= (M_{B,C}^T + M_{B,C}^S) \mathbf{x}$$
Substitution

Since the matrices $M_{B,C}^{T+S}$ and $M_{B,C}^{T} + M_{B,C}^{S}$ have equal matrix-vector products for *every* vector in \mathbb{C}^{n} , by Theorem EMMVP [466] they are equal matrices. (Now would be a good time to double-back and study the proof of Theorem EMMVP [466]. You did promise to come back to this theorem sometime, didn't you?)

Theorem MRMLT Matrix Representation of a Multiple of a Linear Transformation Suppose that $T: U \to V$ is a linear transformation, $\alpha \in \mathbb{C}$, B is a basis of U and C is a basis

 \square

of V. Then

$$M_{B,C}^{\alpha T} = \alpha M_{B,C}^T$$

Proof Let \mathbf{x} be any vector in \mathbb{C}^n . Define $\mathbf{u} \in U$ by $\mathbf{u} = \rho_B^{-1}(\mathbf{x})$, so $\mathbf{x} = \rho_B(\mathbf{u})$. Then,

$M_{B,C}^{\alpha T} \mathbf{x} = M_{B,C}^{\alpha T} \rho_B \left(\mathbf{u} \right)$	Substitution
$=\rho_{C}\left(\left(\alpha T\right)\left(\mathbf{u}\right)\right)$	Theorem FTMR [1293]
$=\rho_{C}\left(\alpha T\left(\mathbf{u}\right)\right)$	Definition LTSM [1108]
$=\alpha\rho_{C}\left(T\left(\mathbf{u}\right)\right)$	Definition LT $[1072]$
$=\alpha\left(M_{B,C}^{T}\rho_{B}\left(\mathbf{u}\right)\right)$	Theorem FTMR [1293]
$=\left(\alpha M_{B,C}^{T}\right)\rho_{B}\left(\mathbf{u}\right)$	Theorem MMSMM [477]
$= \left(lpha M_{B,C}^T ight) \mathbf{x}$	Substitution

Since the matrices $M_{B,C}^{\alpha T}$ and $\alpha M_{B,C}^{T}$ have equal matrix-vector products for *every* vector in \mathbb{C}^{n} , by Theorem EMMVP [466] they are equal matrices.

The vector space of all linear transformations from U to V is now isomorphic to the vector space of all $m \times n$ matrices.

Theorem MRCLT Matrix Representation of a Composition of Linear Transformations Suppose that $T: U \to V$ and $S: V \to W$ are linear transformations, B is a basis of U, C is

a basis of V, and D is a basis of W. Then

$$M_{B,D}^{S \circ T} = M_{C,D}^S M_{B,C}^T$$

Proof Let \mathbf{x} be any vector in \mathbb{C}^n . Define $\mathbf{u} \in U$ by $\mathbf{u} = \rho_B^{-1}(\mathbf{x})$, so $\mathbf{x} = \rho_B(\mathbf{u})$. Then,

$$M_{B,D}^{S \circ T} \mathbf{x} = M_{B,D}^{S \circ T} \rho_B (\mathbf{u})$$
Substitution
$$= \rho_D ((S \circ T) (\mathbf{u}))$$
Theorem FTMR [1293]
$$= \rho_D (S (T (\mathbf{u})))$$
Definition LTC [1111]
$$= M_{C,D}^S \rho_C (T (\mathbf{u}))$$
Theorem FTMR [1293]
$$= M_{C,D}^S (M_{B,C}^T \rho_B (\mathbf{u}))$$
Theorem FTMR [1293]
$$= (M_{C,D}^S M_{B,C}^T) \rho_B (\mathbf{u})$$
Theorem MMA [478]
$$= (M_{C,D}^S M_{B,C}^T) \mathbf{x}$$
Substitution

Since the matrices $M_{B,D}^{S \circ T}$ and $M_{C,D}^S M_{B,C}^T$ have equal matrix-vector products for *every* vector in \mathbb{C}^n , by Theorem EMMVP [466] they are equal matrices.

This is the second great surprise of introductory linear algebra. Matrices are linear transformations (functions, really), and matrix multiplication is function composition! We can form the composition of two linear transformations, then form the matrix representation of the result. Or we can form the matrix representation of each linear transformation separately, then *multiply* the two representations together via Definition MM [468]. In either case, we arrive at the same result.

Example MPMR Matrix product of matrix representations

Consider the two linear transformations,

$$T: \mathbb{C}^2 \to P_2 \quad T\left(\begin{bmatrix}a\\b\end{bmatrix}\right) = (-a+3b) + (2a+4b)x + (a-2b)x^2$$
$$S: P_2 \to M_{22} \quad S\left(a+bx+cx^2\right) = \begin{bmatrix}2a+b+2c & a+4b-c\\-a+3c & 3a+b+2c\end{bmatrix}$$

and bases for \mathbb{C}^2 , P_2 and M_{22} (respectively),

$$B = \left\{ \begin{bmatrix} 3\\1 \end{bmatrix}, \begin{bmatrix} 2\\1 \end{bmatrix} \right\}$$

$$C = \left\{ 1 - 2x + x^2, -1 + 3x, 2x + 3x^2 \right\}$$

$$D = \left\{ \begin{bmatrix} 1 & -2\\1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & -1\\1 & 2 \end{bmatrix}, \begin{bmatrix} -1 & 2\\0 & 0 \end{bmatrix}, \begin{bmatrix} 2 & -3\\2 & 2 \end{bmatrix} \right\}$$

Begin by computing the new linear transformation that is the composition of T and S(Definition LTC [1111], Theorem CLTLT [1111]), $(S \circ T) : \mathbb{C}^2 \to M_{22}$,

$$(S \circ T) \left(\begin{bmatrix} a \\ b \end{bmatrix} \right) = S \left(T \left(\begin{bmatrix} a \\ b \end{bmatrix} \right) \right)$$
$$= S \left((-a+3b) + (2a+4b)x + (a-2b)x^2 \right)$$

$$= \begin{bmatrix} 2(-a+3b) + (2a+4b) + 2(a-2b) & (-a+3b) + 4(2a+4b) - (a-2b) \\ -(-a+3b) + 3(a-2b) & 3(-a+3b) + (2a+4b) + 2(a-2b) \end{bmatrix}$$
$$= \begin{bmatrix} 2a+6b & 6a+21b \\ 4a-9b & a+9b \end{bmatrix}$$

Now compute the matrix representations (Definition MR [1286]) for each of these three linear transformations $(T, S, S \circ T)$, relative to the appropriate bases. First for T,

$$\rho_C \left(T \left(\begin{bmatrix} 3\\1 \end{bmatrix} \right) \right) = \rho_C \left(10x + x^2 \right)$$

$$= \rho_C \left(28(1 - 2x + x^2) + 28(-1 + 3x) + (-9)(2x + 3x^2) \right) = \begin{bmatrix} 28\\28\\-9 \end{bmatrix}$$

$$\rho_C \left(T \left(\begin{bmatrix} 2\\1 \end{bmatrix} \right) \right) = \rho_C \left(1 + 8x \right)$$

$$= \rho_C \left(33(1 - 2x + x^2) + 32(-1 + 3x) + (-11)(2x + 3x^2) \right) = \begin{bmatrix} 33\\32\\-11 \end{bmatrix}$$

So we have the matrix representation of T,

$$M_{B,C}^{T} = \begin{bmatrix} 28 & 33\\ 28 & 32\\ -9 & -11 \end{bmatrix}$$

Now, a representation of S,

$$\rho_D \left(S \left(1 - 2x + x^2 \right) \right) = \rho_D \left(\begin{bmatrix} 2 & -8 \\ 2 & 3 \end{bmatrix} \right)$$

$$= \rho_D \left(\left(-11 \right) \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix} + \left(-21 \right) \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} + 0 \begin{bmatrix} -1 & 2 \\ 0 & 0 \end{bmatrix} + \left(17 \right) \begin{bmatrix} 2 & -3 \\ 2 & 2 \end{bmatrix} \right)$$

$$= \begin{bmatrix} -11 \\ -21 \\ 0 \\ 17 \end{bmatrix}$$

$$\rho_D \left(S \left(-1 + 3x \right) \right) = \rho_D \left(\begin{bmatrix} 1 & 11 \\ 1 & 0 \end{bmatrix} \right)$$

$$= \rho_D \left(26 \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix} + 51 \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} + 0 \begin{bmatrix} -1 & 2 \\ 0 & 0 \end{bmatrix} + \left(-38 \right) \begin{bmatrix} 2 & -3 \\ 2 & 2 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 26 \\ 51 \\ 0 \\ -38 \end{bmatrix}$$

$$\rho_D \left(S \left(2x + 3x^2 \right) \right) = \rho_D \left(\begin{bmatrix} 8 & 5 \\ 9 & 8 \end{bmatrix} \right) \\= \rho_D \left(34 \begin{bmatrix} 1 & -2 \\ 1 & -1 \end{bmatrix} + 67 \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} + 1 \begin{bmatrix} -1 & 2 \\ 0 & 0 \end{bmatrix} + (-46) \begin{bmatrix} 2 & -3 \\ 2 & 2 \end{bmatrix} \right) \\= \begin{bmatrix} 34 \\ 67 \\ 1 \\ -46 \end{bmatrix}$$

So we have the matrix representation of S,

$$M_{C,D}^{S} = \begin{bmatrix} -11 & 26 & 34 \\ -21 & 51 & 67 \\ 0 & 0 & 1 \\ 17 & -38 & -46 \end{bmatrix}$$

Finally, a representation of $S \circ T$,

$$\rho_D\left(\left(S \circ T\right) \begin{pmatrix} \begin{bmatrix} 3\\1 \end{bmatrix} \right) \right) = \rho_D\left(\begin{bmatrix} 12 & 39\\3 & 12 \end{bmatrix} \right)$$
$$= \rho_D\left(114 \begin{bmatrix} 1 & -2\\1 & -1 \end{bmatrix} + 237 \begin{bmatrix} 1 & -1\\1 & 2 \end{bmatrix} + (-9) \begin{bmatrix} -1 & 2\\0 & 0 \end{bmatrix} + (-174) \begin{bmatrix} 2 & -3\\2 & 2 \end{bmatrix} \right)$$

$$\begin{aligned}
&= \begin{bmatrix} 114\\237\\-9\\-174 \end{bmatrix} \\
\rho_D\left((S \circ T)\left(\begin{bmatrix}2\\1\end{bmatrix}\right)\right) &= \rho_D\left(\begin{bmatrix}10&33\\-1&11\end{bmatrix}\right) \\
&= \rho_D\left(95\begin{bmatrix}1&-2\\1&-1\end{bmatrix}+202\begin{bmatrix}1&-1\\1&2\end{bmatrix}+(-11)\begin{bmatrix}-1&2\\0&0\end{bmatrix}+(-149)\begin{bmatrix}2&-3\\2&2\end{bmatrix}\right) \\
&= \begin{bmatrix} 95\\202\\-11\\-149 \end{bmatrix}
\end{aligned}$$

So we have the matrix representation of $S \circ T$,

$$M_{B,D}^{S \circ T} = \begin{bmatrix} 114 & 95\\ 237 & 202\\ -9 & -11\\ -174 & -149 \end{bmatrix}$$

Now, we are all set to verify the conclusion of Theorem MRCLT [1302],

$$M_{C,D}^{S}M_{B,C}^{T} = \begin{bmatrix} -11 & 26 & 34 \\ -21 & 51 & 67 \\ 0 & 0 & 1 \\ 17 & -38 & -46 \end{bmatrix} \begin{bmatrix} 28 & 33 \\ 28 & 32 \\ -9 & -11 \end{bmatrix}$$
$$= \begin{bmatrix} 114 & 95 \\ 237 & 202 \\ -9 & -11 \\ -174 & -149 \end{bmatrix}$$
$$= M_{B,D}^{S \circ T}$$

We have intentionally used non-standard bases. If you were to choose "nice" bases for the three vector spaces, then the result of the theorem might be rather transparent. But this would still be a worthwhile exercise — give it a go. \boxtimes

A diagram, similar to ones we have seen earlier, might make the importance of this theorem clearer,

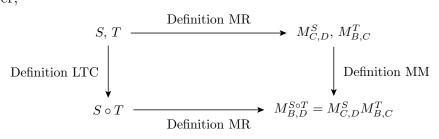


Diagram MRCLT. Matrix Representation and Composition of Linear Transformations

One of our goals in the first part of this book is to make the definition of matrix multiplication (Definition MVP [461], Definition MM [468]) seem as natural as possible. However, many of us are brought up with an entry-by-entry description of matrix multiplication (Theorem EMP [471]) as the *definition* of matrix multiplication, and then theorems about columns of matrices and linear combinations follow from that definition. With this unmotivated definition, the realization that matrix multiplication is function composition is quite remarkable. It is an interesting exercise to begin with the question, "What is the matrix representation of the composition of two linear transformations?" and then, without using any theorems about matrix multiplication, finally arrive at the entry-by-entry description of matrix multiplication. Try it yourself (Exercise MR.T80 [1336]).

Subsection PMR Properties of Matrix Representations

It will not be a surprise to discover that the kernel and range of a linear transformation are closely related to the null space and column space of the transformation's matrix representation. Perhaps this idea has been bouncing around in your head already, even before seeing the definition of a matrix representation. However, with a formal definition of a matrix representation (Definition MR [1286]), and a fundamental theorem to go with it (Theorem

FTMR [1293]) we can be formal about the relationship, using the idea of isomorphic vector spaces (Definition IVS [1230]). Here are the twin theorems.

Theorem KNSI

Kernel and Null Space Isomorphism

Suppose that $T: U \to V$ is a linear transformation, B is a basis for U of size n, and C is a basis for V. Then the kernel of T is isomorphic to the null space of $M_{B,C}^T$,

$$\mathcal{K}(T) \cong \mathcal{N}\big(M_{B,C}^T\big)$$

Proof To establish that two vector spaces are isomorphic, we must find an isomorphism between them, an invertible linear transformation (Definition IVS [1230]). The kernel of the linear transformation T, $\mathcal{K}(T)$, is a subspace of U, while the null space of the matrix representation, $\mathcal{N}(M_{B,C}^T)$ is a subspace of \mathbb{C}^n . The function ρ_B is defined as a function from U to \mathbb{C}^n , but we can just as well employ the definition of ρ_B as a function from $\mathcal{K}(T)$ to $\mathcal{N}(M_{B,C}^T)$.

We must first insure that if we choose an input for ρ_B from $\mathcal{K}(T)$ that then the output will be an element of $\mathcal{N}(M_{B,C}^T)$. So suppose that $\mathbf{u} \in \mathcal{K}(T)$. Then

$M_{B,C}^{T}\rho_{B}\left(\mathbf{u}\right)=\rho_{C}\left(T\left(\mathbf{u}\right)\right)$	Theorem FTMR [1293]
$= ho_{C}\left(0 ight)$	Definition KLT [1138]
= 0	Theorem LTTZZ [1080]

This says that $\rho_B(\mathbf{u}) \in \mathcal{N}(M_{B,C}^T)$, as desired.

The restriction in the size of the domain and codomain ρ_B will not affect the fact that ρ_B is a linear transformation (Theorem VRLT [1262]), nor will it affect the fact that ρ_B is injective (Theorem VRI [1270]). Something must be done though to verify that ρ_B is surjective. To this end, appeal to the definition of surjective (Definition SLT [1169]), and suppose that we have an element of the codomain, $\mathbf{x} \in \mathcal{N}(M_{B,C}^T) \subseteq \mathbb{C}^n$ and we wish to find an element of the domain with \mathbf{x} as its image. We now show that the desired element of the domain is $\mathbf{u} = \rho_B^{-1}(\mathbf{x})$. First, verify that $\mathbf{u} \in \mathcal{K}(T)$,

$$T (\mathbf{u}) = T \left(\rho_B^{-1}(\mathbf{x})\right)$$

$$= \rho_C^{-1} \left(M_{B,C}^T \left(\rho_B \left(\rho_B^{-1}(\mathbf{x})\right)\right)\right) \qquad \text{The}$$

$$= \rho_C^{-1} \left(M_{B,C}^T \left(I_{\mathbb{C}^n}(\mathbf{x})\right)\right) \qquad \text{Def}$$

$$= \rho_C^{-1} \left(M_{B,C}^T \mathbf{x}\right) \qquad \text{Def}$$

$$= \rho_C^{-1} \left(\mathbf{0}_{\mathbb{C}^n}\right) \qquad \text{Def}$$

$$= \mathbf{0}_V \qquad \text{The}$$

Theorem FTMR [1293] Definition IVLT [1214] Definition IDLT [1213] Definition KLT [1138] Theorem LTTZZ [1080]

Second, verify that the proposed isomorphism, ρ_B , takes **u** to **x**,

$$\rho_B \left(\mathbf{u} \right) = \rho_B \left(\rho_B^{-1} \left(\mathbf{x} \right) \right)$$
Substitution
$$= I_{\mathbb{C}^n} \left(\mathbf{x} \right)$$
Definition IVLT [1214]

 $= \mathbf{x}$

Definition IDLT [1213]

With ρ_B demonstrated to be an injective and surjective linear transformation from $\mathcal{K}(T)$ to $\mathcal{N}(M_{B,C}^T)$, Theorem ILTIS [1221] tells us ρ_B is invertible, and so by Definition IVS [1230], we say $\mathcal{K}(T)$ and $\mathcal{N}(M_{B,C}^T)$ are isomorphic.

Example KVMR Kernel via matrix representation

Consider the kernel of the linear transformation

$$T: M_{22} \to P_2, \quad T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (2a - b + c - 5d) + (a + 4b + 5c + 2d)x + (3a - 2b + c - 8d)x^2$$

We will begin with a matrix representation of T relative to the bases for M_{22} and P_2 (respectively),

$$B = \left\{ \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ -1 & -4 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} 2 & 5 \\ -2 & -4 \end{bmatrix} \right\}$$
$$C = \left\{ 1 + x + x^2, 2 + 3x, -1 - 2x^2 \right\}$$

Then,

$$\rho_C \left(T \left(\begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} \right) \right) = \rho_C \left(4 + 2x + 6x^2 \right)$$
$$= \rho_C \left(2(1 + x + x^2) + 0(2 + 3x) + (-2)(-1 - 2x^2) \right)$$

$$= \begin{bmatrix} 2\\0\\-2 \end{bmatrix}$$

$$\rho_C \left(T\left(\begin{bmatrix} 1 & 3\\-1 & -4 \end{bmatrix} \right) \right) = \rho_C \left(18 + 28x^2 \right)$$

$$= \rho_C \left((-24)(1 + x + x^2) + 8(2 + 3x) + (-26)(-1 - 2x^2) \right)$$

$$= \begin{bmatrix} -24\\8\\-26 \end{bmatrix}$$

$$\rho_C \left(T\left(\begin{bmatrix} 1 & 2\\0 & -2 \end{bmatrix} \right) \right) = \rho_C \left(10 + 5x + 15x^2 \right)$$

$$= \rho_C \left(5(1 + x + x^2) + 0(2 + 3x) + (-5)(-1 - 2x^2) \right)$$

$$= \begin{bmatrix} 5\\0\\-5 \end{bmatrix}$$

$$\rho_C \left(T\left(\begin{bmatrix} 2 & 5\\-2 & -4 \end{bmatrix} \right) \right) = \rho_C \left(17 + 4x + 26x^2 \right)$$

$$= \rho_C \left((-8)(1 + x + x^2) + (4)(2 + 3x) + (-17)(-1 - 2x^2) \right)$$

$$= \begin{bmatrix} -8\\4\\-17 \end{bmatrix}$$

So the matrix representation of T (relative to B and C) is

$$M_{B,C}^{T} = \begin{bmatrix} 2 & -24 & 5 & -8 \\ 0 & 8 & 0 & 4 \\ -2 & -26 & -5 & -17 \end{bmatrix}$$

We know from Theorem KNSI [1311] that the kernel of the linear transformation T is isomorphic to the null space of the matrix representation $M_{B,C}^T$ and by studying the proof of Theorem KNSI [1311] we learn that ρ_B is an isomorphism between these null spaces. Rather than trying to compute the kernel of T using definitions and techniques from Chapter LT [1071] we will instead analyze the null space of $M_{B,C}^T$ using techniques from way back in Chapter V [197]. First row-reduce $M_{B,C}^T$,

$$\begin{bmatrix} 2 & -24 & 5 & -8\\ 0 & 8 & 0 & 4\\ -2 & -26 & -5 & -17 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & \frac{5}{2} & 2\\ 0 & \boxed{1} & 0 & \frac{1}{2}\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So, by Theorem BNS [334], a basis for $\mathcal{N}(M_{B,C}^T)$ is

$$\left\langle \left\{ \begin{bmatrix} -\frac{5}{2} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$

We can now convert this basis of $\mathcal{N}(M_{B,C}^T)$ into a basis of $\mathcal{K}(T)$ by applying ρ_B^{-1} to each element of the basis,

$$\begin{split} \rho_B^{-1} \left(\begin{bmatrix} -\frac{5}{2} \\ 0 \\ 1 \\ 0 \end{bmatrix} \right) &= (-\frac{5}{2}) \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} + 0 \begin{bmatrix} 1 & 3 \\ -1 & -4 \end{bmatrix} + 1 \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} + 0 \begin{bmatrix} 2 & 5 \\ -2 & -4 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{3}{2} & -3 \\ \frac{5}{2} & \frac{1}{2} \end{bmatrix} \\ \rho_B^{-1} \left(\begin{bmatrix} -2 \\ -\frac{1}{2} \\ 0 \\ 1 \end{bmatrix} \right) &= (-2) \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} + (-\frac{1}{2}) \begin{bmatrix} 1 & 3 \\ -1 & -4 \end{bmatrix} + 0 \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} + 1 \begin{bmatrix} 2 & 5 \\ -2 & -4 \end{bmatrix} \\ &= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix} \end{split}$$

So the set

$$\left\{ \begin{bmatrix} -\frac{3}{2} & -3\\ \frac{5}{2} & \frac{1}{2} \end{bmatrix}, \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2}\\ \frac{1}{2} & 0 \end{bmatrix} \right\}$$

is a basis for $\mathcal{K}(T)$ Just for fun, you might evaluate T with each of these two basis vectors and verify that the output is the zero polynomial (Exercise MR.C10 [1331]).

An entirely similar result applies to the range of a linear transformation and the column space of a matrix representation of the linear transformation.

Theorem RCSI Range and Column Space Isomorphism

Suppose that $T: U \to V$ is a linear transformation, B is a basis for U of size n, and C is a basis for V of size m. Then the range of T is isomorphic to the column space of $M_{B,C}^T$,

$$\mathcal{R}(T) \cong \mathcal{C}(M_{B,C}^T)$$

 \square

Proof To establish that two vector spaces are isomorphic, we must find an isomorphism between them, an invertible linear transformation (Definition IVS [1230]). The range of the linear transformation T, $\mathcal{R}(T)$, is a subspace of V, while the column space of the matrix representation, $\mathcal{C}(M_{B,C}^T)$ is a subspace of \mathbb{C}^m . The function ρ_C is defined as a function from V to \mathbb{C}^m , but we can just as well employ the definition of ρ_C as a function from $\mathcal{R}(T)$ to $\mathcal{C}(M_{B,C}^T)$.

We must first insure that if we choose an input for ρ_C from $\mathcal{R}(T)$ that then the output will be an element of $\mathcal{C}(M_{B,C}^T)$. So suppose that $\mathbf{v} \in \mathcal{R}(T)$. Then there is a vector $\mathbf{u} \in U$, such that $T(\mathbf{u}) = \mathbf{v}$. Consider

$$M_{B,C}^{T}\rho_{B}\left(\mathbf{u}\right) = \rho_{C}\left(T\left(\mathbf{u}\right)\right) \qquad \text{Theorem FTMR [1293]} \\ = \rho_{C}\left(\mathbf{v}\right) \qquad \text{Definition RLT [1178]}$$

This says that $\rho_C(\mathbf{v}) \in \mathcal{C}(M_{B,C}^T)$, as desired.

The restriction in the size of the domain and codomain will not affect the fact that ρ_C is a linear transformation (Theorem VRLT [1262]), nor will it affect the fact that ρ_C

is injective (Theorem VRI [1270]). Something must be done though to verify that ρ_C is surjective. This all gets a bit confusing, since the domain of our isomorphism is the range of the linear transformation, so think about your objects as you go. To establish that ρ_C is surjective, appeal to the definition of a surjective linear transformation (Definition SLT [1169]), and suppose that we have an element of the codomain, $\mathbf{y} \in \mathcal{C}(M_{B,C}^T) \subseteq \mathbb{C}^m$ and we wish to find an element of the domain with \mathbf{y} as its image. Since $\mathbf{y} \in \mathcal{C}(M_{B,C}^T)$, there exists a vector, $\mathbf{x} \in \mathbb{C}^n$ with $M_{B,C}^T \mathbf{x} = \mathbf{y}$. We now show that the desired element of the domain is $\mathbf{v} = \rho_C^{-1}(\mathbf{y})$. First, verify that $\mathbf{v} \in \mathcal{R}(T)$ by applying T to $\mathbf{u} = \rho_B^{-1}(\mathbf{x})$,

$$T \left(\mathbf{u} \right) = T \left(\rho_B^{-1} \left(\mathbf{x} \right) \right)$$
$$= \rho_C^{-1} \left(M_{B,C}^T \left(\rho_B \left(\rho_B^{-1} \left(\mathbf{x} \right) \right) \right) \right)$$
$$= \rho_C^{-1} \left(M_{B,C}^T \left(I_{\mathbb{C}^n} \left(\mathbf{x} \right) \right) \right)$$
$$= \rho_C^{-1} \left(M_{B,C}^T \mathbf{x} \right)$$
$$= \rho_C^{-1} \left(\mathbf{y} \right)$$
$$= \mathbf{v}$$

Theorem FTMR [1293] Definition IVLT [1214] Definition IDLT [1213] Definition CSM [560] Substitution

Second, verify that the proposed isomorphism, ρ_C , takes **v** to **y**,

$$\rho_{C} (\mathbf{v}) = \rho_{C} \left(\rho_{C}^{-1} (\mathbf{y}) \right) \qquad \text{Substitution} \\
= I_{\mathbb{C}^{m}} (\mathbf{y}) \qquad \text{Definition IVLT [1214]} \\
= \mathbf{y} \qquad \text{Definition IDLT [1213]}$$

With ρ_C demonstrated to be an injective and surjective linear transformation from $\mathcal{R}(T)$ to $\mathcal{C}(M_{B,C}^T)$, Theorem ILTIS [1221] tells us ρ_C is invertible, and so by Definition IVS [1230], we say $\mathcal{R}(T)$ and $\mathcal{C}(M_{B,C}^T)$ are isomorphic.

Example RVMR Range via matrix representation

In this example, we will recycle the linear transformation T and the bases B and C of Example KVMR [1313] but now we will compute the range of T,

$$T: M_{22} \to P_2, \quad T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (2a - b + c - 5d) + (a + 4b + 5b + 2d)x + (3a - 2b + c - 8d)x^2$$

With bases B and C,

$$B = \left\{ \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ -1 & -4 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} 2 & 5 \\ -2 & -4 \end{bmatrix} \right\}$$
$$C = \left\{ 1 + x + x^2, 2 + 3x, -1 - 2x^2 \right\}$$

we obtain the matrix representation

$$M_{B,C}^{T} = \begin{bmatrix} 2 & -24 & 5 & -8 \\ 0 & 8 & 0 & 4 \\ -2 & -26 & -5 & -17 \end{bmatrix}$$

We know from Theorem RCSI [1317] that the range of the linear transformation T is isomorphic to the column space of the matrix representation $M_{B,C}^T$ and by studying the proof of

Theorem RCSI [1317] we learn that ρ_C is an isomorphism between these subspaces. Notice that since the range is a subspace of the codomain, we will employ ρ_C as the isomorphism, rather than ρ_B , which was the correct choice for an isomorphism between the null spaces of Example KVMR [1313].

Rather than trying to compute the range of T using definitions and techniques from Chapter LT [1071] we will instead analyze the column space of $M_{B,C}^T$ using techniques from way back in Chapter M [425]. First row-reduce $(M_{B,C}^T)^t$,

$$\begin{bmatrix} 2 & 0 & -2 \\ -24 & 8 & -26 \\ 5 & 0 & -5 \\ -8 & 4 & -17 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -\frac{25}{4} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Now employ Theorem CSRST [586] and Theorem BRS [583] (there are other methods we could choose here to compute the column space, such as Theorem BCS [569]) to obtain the basis for $\mathcal{C}(M_{B,C}^T)$,

$$\left\{ \begin{bmatrix} 1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\1\\-\frac{25}{4} \end{bmatrix} \right\}$$

We can now convert this basis of $\mathcal{C}(M_{B,C}^T)$ into a basis of $\mathcal{R}(T)$ by applying ρ_C^{-1} to each

element of the basis,

$$\rho_C^{-1} \left(\begin{bmatrix} 1\\0\\-1 \end{bmatrix} \right) = (1+x+x^2) - (-1-2x^2) = 2+x+3x^2$$
$$\rho_C^{-1} \left(\begin{bmatrix} 0\\1\\-\frac{25}{4} \end{bmatrix} \right) = (2+3x) - \frac{25}{4}(-1-2x^2) = \frac{33}{4} + 3x + \frac{31}{2}x^2$$

So the set

$$\left\{2+3x+3x^2, \frac{33}{4}+3x+\frac{31}{2}x^2\right\}$$

is a basis for $\mathcal{R}(T)$.

Theorem KNSI [1311] and Theorem RCSI [1317] can be viewed as further formal evidence for the Coordinatization Principle [1279], though they are not direct consequences.

Subsection IVLT Invertible Linear Transformations

We have seen, both in theorems and in examples, that questions about linear transformations are often equivalent to questions about matrices. It is the matrix representation of a linear

 \boxtimes

transformation that makes this idea precise. Here's our final theorem that solidifies this connection.

Theorem IMR Invertible Matrix Representations

Suppose that $T: U \to V$ is a linear transformation, B is a basis for U and C is a basis for V. Then T is an invertible linear transformation if and only if the matrix representation of T relative to B and C, $M_{B,C}^{T}$ is an invertible matrix. When T is invertible,

$$M_{C,B}^{T^{-1}} = \left(M_{B,C}^T\right)^{-1}$$

Proof (\Leftarrow) Suppose *T* is invertible, so the inverse linear transformation $T^{-1}: V \to U$ exists (Definition IVLT [1214]). Both linear transformations have matrix representations relative to the bases of *U* and *V*, namely $M_{B,C}^T$ and $M_{C,B}^{T^{-1}}$ (Definition MR [1286]). Then

$$M_{C,B}^{T^{-1}}M_{B,C}^{T} = M_{B,B}^{T^{-1}\circ T}$$

$$= M_{B,B}^{I_{U}}$$

$$= [\rho_{B}(I_{U}(\mathbf{u}_{1}))| \rho_{B}(I_{U}(\mathbf{u}_{2}))| \dots |\rho_{B}(I_{U}(\mathbf{u}_{n}))]$$

$$= [\rho_{B}(\mathbf{u}_{1})| \rho_{B}(\mathbf{u}_{2})| \rho_{B}(\mathbf{u}_{3})| \dots |\rho_{B}(\mathbf{u}_{n})]$$

$$= [\mathbf{e}_{1}|\mathbf{e}_{2}|\mathbf{e}_{3}| \dots |\mathbf{e}_{n}]$$

$$= I_{n}$$
Theorem MRCLT [1302]
Definition IVLT [1214]
Definition MR [1286]
Definition IDLT [1213]
Definition VR [1261]
Definition IM [175]

and

$$\begin{split} M_{B,C}^{T} M_{C,B}^{T^{-1}} &= M_{C,C}^{T \circ T^{-1}} & \text{Theoder} \\ &= M_{C,C}^{I_{V}} & \text{Define } \\ &= \left[\rho_{C} \left(I_{V} \left(\mathbf{v}_{1}\right)\right)\right| \rho_{C} \left(I_{V} \left(\mathbf{v}_{2}\right)\right)\right| \dots \left|\rho_{C} \left(I_{V} \left(\mathbf{v}_{n}\right)\right)\right] & \text{Define } \\ &= \left[\rho_{C} \left(\mathbf{v}_{1}\right)\right| \rho_{C} \left(\mathbf{v}_{2}\right)\right| \rho_{C} \left(\mathbf{v}_{3}\right)\right| \dots \left|\rho_{C} \left(\mathbf{v}_{n}\right)\right] & \text{Define } \\ &= \left[\mathbf{e}_{1} |\mathbf{e}_{2}| \mathbf{e}_{3}| \dots |\mathbf{e}_{n}\right] & \text{Define } \\ &= I_{n} & \text{Define } \end{split}$$

Theorem MRCLT [1302] Definition IVLT [1214] Definition MR [1286] Definition IDLT [1213] Definition VR [1261] Definition IM [175]

These two equations show that $M_{B,C}^T$ and $M_{C,B}^{T^{-1}}$ are inverse matrices (Definition MI [506]) and establish that when T is invertible, then $M_{C,B}^{T^{-1}} = (M_{B,C}^T)^{-1}$.

(\Leftarrow) Suppose now that $M_{B,C}^T$ is an invertible matrix and hence nonsingular (Theorem NI [542]). We compute the nullity of T,

$$n(T) = \dim (\mathcal{K}(T)) \qquad \text{Definition KLT [1138]} \\ = \dim (\mathcal{N}(M_{B,C}^T)) \qquad \text{Theorem KNSI [1311]} \\ = n(M_{B,C}^T) \qquad \text{Definition NOM [829]} \\ = 0 \qquad \text{Theorem RNNM [833]}$$

So the kernel of T is trivial, and by Theorem KILT [1144], T is injective.

We now compute the rank of T,

$$r(T) = \dim (\mathcal{R}(T))$$
 De

Definition RLT [1178]

$= \dim \left(\mathcal{C} \left(M_{B,C}^T \right) \right)$	Theorem RCSI [1317]
$= r\left(M_{B,C}^{T}\right)$	Definition ROM [829]
$=\dim\left(V\right)$	Theorem RNNM [833]

Since the dimension of the range of T equals the dimension of the codomain V, by Theorem EDYES [858], $\mathcal{R}(T) = V$. Which says that T is surjective by Theorem RSLT [1185]. Because T is both injective and surjective, by Theorem ILTIS [1221], T is invertible.

By now, the connections between matrices and linear transformations should be starting to become more transparent, and you may have already recognized the invertibility of a matrix as being tantamount to the invertibility of the associated matrix representation. The next example shows how to apply this theorem to the problem of actually building a formula for the inverse of an invertible linear transformation.

Example ILTVR

Inverse of a linear transformation via a representation

Consider the linear transformation

$$R: P_3 \to M_{22}, \quad R\left(a + bx + cx^2 + x^3\right) = \begin{bmatrix} a + b - c + 2d & 2a + 3b - 2c + 3d \\ a + b + 2d & -a + b + 2c - 5d \end{bmatrix}$$

If we wish to quickly find a formula for the inverse of R (presuming it exists), then choosing "nice" bases will work best. So build a matrix representation of R relative to the bases B and C,

$$B = \left\{1, \, x, \, x^2, \, x^3\right\}$$

$$C = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Then,

$$\rho_C(R(1)) = \rho_C\left(\begin{bmatrix} 1 & 2\\ 1 & -1 \end{bmatrix}\right) = \begin{bmatrix} 1\\ 2\\ 1\\ -1 \end{bmatrix}$$
$$\rho_C(R(x)) = \rho_C\left(\begin{bmatrix} 1 & 3\\ 1 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1\\ 3\\ 1\\ 1\\ 1 \end{bmatrix}$$
$$\rho_C(R(x^2)) = \rho_C\left(\begin{bmatrix} -1 & -2\\ 0 & 2 \end{bmatrix}\right) = \begin{bmatrix} -1\\ -2\\ 0\\ 2\\ 0\\ 2 \end{bmatrix}$$
$$\rho_C(R(x^3)) = \rho_C\left(\begin{bmatrix} 2 & 3\\ 2 & -5 \end{bmatrix}\right) = \begin{bmatrix} 2\\ 3\\ 2\\ -5 \end{bmatrix}$$

So a representation of R is

$$M_{B,C}^{R} = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 2 & 3 & -2 & 3 \\ 1 & 1 & 0 & 2 \\ -1 & 1 & 2 & -5 \end{bmatrix}$$

The matrix $M_{B,C}^R$ is invertible (as you can check) so we know for sure that R is invertible by Theorem IMR [1322]. Furthermore,

$$M_{C,B}^{R^{-1}} = \left(M_{B,C}^{R}\right)^{-1} = \begin{bmatrix} 1 & 1 & -1 & 2\\ 2 & 3 & -2 & 3\\ 1 & 1 & 0 & 2\\ -1 & 1 & 2 & -5 \end{bmatrix}^{-1} = \begin{bmatrix} 20 & -7 & -2 & 3\\ -8 & 3 & 1 & -1\\ -1 & 0 & 1 & 0\\ -6 & 2 & 1 & -1 \end{bmatrix}$$

We can use this representation of the inverse linear transformation, in concert with Theorem FTMR [1293], to determine an explicit formula for the inverse itself,

$$R^{-1}\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right) = \rho_B^{-1}\left(M_{C,B}^{R^{-1}}\rho_C\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right)\right)$$
 Theorem FTMR [1293]
$$= \rho_B^{-1}\left(\left(M_{B,C}^R\right)^{-1}\rho_C\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right)\right)$$
 Theorem IMR [1322]
$$= \rho_B^{-1}\left(\left(M_{B,C}^R\right)^{-1}\begin{bmatrix}a\\b\\c\\d\end{bmatrix}\right)$$
 Definition VR [1261]

$$= \rho_B^{-1} \left(\begin{bmatrix} 20 & -7 & -2 & 3 \\ -8 & 3 & 1 & -1 \\ -1 & 0 & 1 & 0 \\ -6 & 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \right)$$
 Definition MI [506]
$$= \rho_B^{-1} \left(\begin{bmatrix} 20a - 7b - 2c + 3d \\ -8a + 3b + c - d \\ -a + c \\ -6a + 2b + c - d \end{bmatrix} \right)$$
 Definition MVP [461]
$$= (20a - 7b - 2c + 3d) + (-8a + 3b + c - d)x + (-a + c)x^2 + (-6a + 2b + c - d)x^3$$
 Definition VR [1261]

 \boxtimes

You might look back at Example AIVLT [1214], where we first witnessed the inverse of a linear transformation and recognize that the inverse (S) was built from using the method of Example ILTVR [1324] with a matrix representation of T.

Theorem IMILT

Invertible Matrices, Invertible Linear Transformation

Suppose that A is a square matrix of size n and $T: \mathbb{C}^n \to \mathbb{C}^n$ is the linear transformation defined by $T(\mathbf{x}) = A\mathbf{x}$. Then A is invertible matrix if and only if T is an invertible linear transformation.

Proof Choose bases $B = C = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_n\}$ consisting of the standard unit vectors as a basis of \mathbb{C}^n (Theorem SUVB [779]) and build a matrix representation of T relative to

 ${\cal B}$ and ${\cal C}.$ Then

$$\rho_C \left(T \left(\mathbf{e}_i \right) \right) = \rho_C \left(A \mathbf{e}_i \right)$$
$$= \rho_C \left(\mathbf{A}_i \right)$$
$$= \mathbf{A}_i$$

So then the matrix representation of T, relative to B and C, is simply $M_{B,C}^T = A$. with this observation, the proof becomes a specialization of Theorem IMR [1322],

T is invertible
$$\iff M_{B,C}^T$$
 is invertible $\iff A$ is invertible

This theorem may seem gratuitous. Why state such a special case of Theorem IMR [1322]? Because it adds another condition to our NMEx series of theorems, and in some ways it is the most fundamental expression of what it means for a matrix to be nonsingular — the associated linear transformation is invertible. This is our final update.

Theorem NME9 Nonsingular Matrix Equivalences, Round 9 Suppose that A is a square matrix of size n. The following are equivalent.

- 1. A is nonsingular.
- 2. A row-reduces to the identity matrix.

- 3. The null space of A contains only the zero vector, $\mathcal{N}(A) = \{\mathbf{0}\}.$
- 4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of \mathbf{b} .
- 5. The columns of A are a linearly independent set.
- 6. A is invertible.
- 7. The column space of A is \mathbb{C}^n , $\mathcal{C}(A) = \mathbb{C}^n$.
- 8. The columns of A are a basis for \mathbb{C}^n .
- 9. The rank of A is n, r(A) = n.
- 10. The nullity of A is zero, n(A) = 0.
- 11. The determinant of A is nonzero, $det(A) \neq 0$.
- 12. $\lambda = 0$ is not an eigenvalue of A.
- 13. The linear transformation $T: \mathbb{C}^n \to \mathbb{C}^n$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is invertible.

Proof By Theorem IMILT [1327] the new addition to this list is equivalent to the statement that A is invertible so we can expand Theorem NME8 [1004].

Subsection READ Reading Questions

- 1. Why does Theorem FTMR [1293] deserve the moniker "fundamental"?
- 2. Find the matrix representation, $M_{B,C}^T$ of the linear transformation

$$T: \mathbb{C}^2 \to \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1\\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 - x_2\\ 3x_1 + 2x_2 \end{bmatrix}$$

relative to the bases

$$B = \left\{ \begin{bmatrix} 2\\3 \end{bmatrix}, \begin{bmatrix} -1\\2 \end{bmatrix} \right\} \qquad \qquad C = \left\{ \begin{bmatrix} 1\\0 \end{bmatrix}, \begin{bmatrix} 1\\1 \end{bmatrix} \right\}$$

3. What is the second "surprise," and why is it surprising?

Subsection EXC Exercises

C10 Example KVMR [1313] concludes with a basis for the kernel of the linear transformation T. Compute the value of T for each of these two basis vectors. Did you get what you expected?

Contributed by Robert Beezer

C20 Compute the matrix representation of T relative to the bases B and C.

$$T: P_3 \to \mathbb{C}^3, \quad T(a+bx+cx^2+dx^3) = \begin{bmatrix} 2a-3b+4c-2d\\a+b-c+d\\3a+2c-3d \end{bmatrix}$$
$$B = \{1, x, x^2, x^3\} \qquad C = \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [1337]

C21 Find a matrix representation of the linear transformation T relative to the bases B and C.

$$T: P_2 \to \mathbb{C}^2, \quad T(p(x)) = \begin{bmatrix} p(1) \\ p(3) \end{bmatrix}$$

$$B = \left\{2 - 5x + x^2, 1 + x - x^2, x^2\right\}$$
$$C = \left\{ \begin{bmatrix} 3\\4 \end{bmatrix}, \begin{bmatrix} 2\\3 \end{bmatrix} \right\}$$

Contributed by Robert Beezer Solution [1338]

C22 Let S_{22} be the vector space of 2×2 symmetric matrices. Build the matrix representation of the linear transformation $T: P_2 \to S_{22}$ relative to the bases B and C and then use this matrix representation to compute $T(3 + 5x - 2x^2)$.

$$B = \{1, 1+x, 1+x+x^2\} \qquad C = \left\{ \begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix} \right\}$$
$$T(a+bx+cx^2) = \begin{bmatrix} 2a-b+c & a+3b-c\\ a+3b-c & a-c \end{bmatrix}$$

Contributed by Robert Beezer Solution [1339]

C25 Use a matrix representation to determine if the linear transformation $T: P_3 \rightarrow M_{22}$ surjective.

$$T(a + bx + cx^{2} + dx^{3}) = \begin{bmatrix} -a + 4b + c + 2d & 4a - b + 6c - d\\ a + 5b - 2c + 2d & a + 2c + 5d \end{bmatrix}$$

Contributed by Robert Beezer Solution [1340]

C30 Find bases for the kernel and range of the linear transformation S below.

$$S: M_{22} \to P_2, \quad S\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+2b+5c-4d) + (3a-b+8c+2d)x + (a+b+4c-2d)x^2$$

Contributed by Robert Beezer Solution [1342]

C40 Let S_{22} be the set of 2×2 symmetric matrices. Verify that the linear transformation R is invertible and find R^{-1} .

$$R: S_{22} \to P_2, \quad R\left(\begin{bmatrix} a & b \\ b & c \end{bmatrix}\right) = (a-b) + (2a-3b-2c)x + (a-b+c)x^2$$

Contributed by Robert Beezer Solution [1343]

C41 Prove that the linear transformation S is invertible. Then find a formula for the inverse linear transformation, S^{-1} , by employing a matrix inverse.

$$S: P_1 \to M_{1,2}, \quad S(a+bx) = \begin{bmatrix} 3a+b & 2a+b \end{bmatrix}$$

Contributed by Robert Beezer Solution [1344]

C42 The linear transformation $R: M_{12} \to M_{21}$ is invertible. Use a matrix representation to determine a formula for the inverse linear transformation $R^{-1}: M_{21} \to M_{12}$.

$$R\left(\begin{bmatrix}a & b\end{bmatrix}\right) = \begin{bmatrix}a+3b\\4a+11b\end{bmatrix}$$

Contributed by Robert Beezer Solution [1346]

C50 Use a matrix representation to find a basis for the range of the linear transformation L.

$$L: M_{22} \to P_2, \quad T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+2b+4c+d) + (3a+c-2d)x + (-a+b+3c+3d)x^2$$

Contributed by Robert Beezer Solution [1347]

C51 Use a matrix representation to find a basis for the kernel of the linear transformation L.

$$L: M_{22} \to P_2, \quad T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a+2b+4c+d) + (3a+c-2d)x + (-a+b+3c+3d)x^2$$

Contributed by Robert Beezer

C52 Find a basis for the kernel of the linear transformation $T: P_2 \to M_{22}$.

$$T(a + bx + cx^{2}) = \begin{bmatrix} a + 2b - 2c & 2a + 2b \\ -a + b - 4c & 3a + 2b + 2c \end{bmatrix}$$

Contributed by Robert Beezer Solution [1348]

M20 The linear transformation D performs differentiation on polynomials. Use a matrix representation of D to find the rank and nullity of D.

$$D: P_n \to P_n, \quad D(p(x)) = p'(x)$$

Contributed by Robert Beezer Solution [1350]

M60 Suppose U and V are vector spaces and define a function $Z: U \to V$ by $T(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Then Exercise IVLT.M60 [1248] asks you to formulate the theorem: Z is invertible if and only if $U = \{\mathbf{0}_U\}$ and $V = \{\mathbf{0}_V\}$. What would a matrix representation of Z look like in this case? How does Theorem IMR [1322] read in this case? Contributed by Robert Beezer

M80 In light of Theorem KNSI [1311] and Theorem MRCLT [1302], write a short comparison of Exercise MM.T40 [494] with Exercise ILT.T15 [1158]. Contributed by Robert Beezer

M81 In light of Theorem RCSI [1317] and Theorem MRCLT [1302], write a short comparison of Exercise CRS.T40 [599] with Exercise SLT.T15 [1201]. Contributed by Robert Beezer

M82 In light of Theorem MRCLT [1302] and Theorem IMR [1322], write a short comparison of Theorem SS [520] and Theorem ICLT [1227]. Contributed by Robert Beezer **M83** In light of Theorem MRCLT [1302] and Theorem IMR [1322], write a short comparison of Theorem NPNT [538] and Exercise IVLT.T40 [1249]. Contributed by Robert Beezer

T20 Construct a new solution to Exercise B.T50 [804] along the following outline. From the $n \times n$ matrix A, construct the linear transformation $T: \mathbb{C}^n \to \mathbb{C}^n$, $T(\mathbf{x}) = A\mathbf{x}$. Use Theorem NI [542], Theorem IMILT [1327] and Theorem ILTIS [1221] to translate between the nonsingularity of A and the surjectivity/injectivity of T. Then apply Theorem ILTB [1149] and Theorem SLTB [1192] to connect these properties with bases. Contributed by Robert Beezer Solution [1352]

T60 Create an entirely different proof of Theorem IMILT [1327] that relies on Definition IVLT [1214] to establish the invertibility of T, and that relies on Definition MI [506] to establish the invertibility of A.

Contributed by Robert Beezer

T80 Suppose that $T: U \to V$ and $S: V \to W$ are linear transformations, and that B, C and D are bases for U, V, and W. Using only Definition MR [1286] define matrix representations for T and S. Using these two definitions, and Definition MR [1286], derive a matrix representation for the composition $S \circ T$ in terms of the entries of the matrices $M_{B,C}^T$ and $M_{C,D}^S$. Explain how you would use this result to motivate a definition for matrix multiplication that is strikingly similar to Theorem EMP [471]. Contributed by Robert Beezer Solution [1352]

Subsection SOL Solutions

C20 Contributed by Robert Beezer Statement [1331] Apply Definition MR [1286],

$$\rho_{C}(T(1)) = \rho_{C}\left(\begin{bmatrix}2\\1\\3\end{bmatrix}\right) = \rho_{C}\left(1\begin{bmatrix}1\\0\\0\end{bmatrix} + (-2)\begin{bmatrix}1\\1\\0\end{bmatrix} + 3\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}1\\-2\\3\end{bmatrix}$$

$$\rho_{C}(T(x)) = \rho_{C}\left(\begin{bmatrix}-3\\1\\0\end{bmatrix}\right) = \rho_{C}\left((-4)\begin{bmatrix}1\\0\\0\end{bmatrix} + 1\begin{bmatrix}1\\1\\0\end{bmatrix} + 0\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}-4\\1\\0\end{bmatrix}$$

$$\rho_{C}(T(x^{2})) = \rho_{C}\left(\begin{bmatrix}4\\-1\\2\end{bmatrix}\right) = \rho_{C}\left(5\begin{bmatrix}1\\0\\0\end{bmatrix} + (-3)\begin{bmatrix}1\\1\\0\end{bmatrix} + 2\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}5\\-3\\2\end{bmatrix}$$

$$\rho_{C}(T(x^{3})) = \rho_{C}\left(\begin{bmatrix}-2\\1\\-3\end{bmatrix}\right) = \rho_{C}\left((-3)\begin{bmatrix}1\\0\\0\end{bmatrix} + 4\begin{bmatrix}1\\1\\0\end{bmatrix} + (-3)\begin{bmatrix}1\\1\\1\end{bmatrix}\right) = \begin{bmatrix}-3\\4\\-3\end{bmatrix}$$

These four vectors are the columns of the matrix representation,

$$M_{B,C}^{T} = \begin{bmatrix} 1 & -4 & 5 & -3 \\ -2 & 1 & -3 & 4 \\ 3 & 0 & 2 & -3 \end{bmatrix}$$

C21 Contributed by Robert Beezer Statement [1331] Applying Definition MR [1286],

$$\rho_C \left(T \left(2 - 5x + x^2 \right) \right) = \rho_C \left(\begin{bmatrix} -2 \\ -4 \end{bmatrix} \right) = \rho_C \left(2 \begin{bmatrix} 3 \\ 4 \end{bmatrix} + (-4) \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$$

$$\rho_C \left(T \left(1 + x - x^2 \right) \right) = \rho_C \left(\begin{bmatrix} 1 \\ -5 \end{bmatrix} \right) = \rho_C \left(13 \begin{bmatrix} 3 \\ 4 \end{bmatrix} + (-19) \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right) = \begin{bmatrix} 13 \\ -19 \end{bmatrix}$$

$$\rho_C \left(T \left(x^2 \right) \right) = \rho_C \left(\begin{bmatrix} 1 \\ 9 \end{bmatrix} \right) = \rho_C \left((-15) \begin{bmatrix} 3 \\ 4 \end{bmatrix} + 23 \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right) = \begin{bmatrix} -15 \\ 23 \end{bmatrix}$$

So the resulting matrix representation is

$$M_{B,C}^T = \begin{bmatrix} 2 & 13 & -15 \\ -4 & -19 & 23 \end{bmatrix}$$

C22 Contributed by Robert Beezer Statement [1332]

Input to T the vectors of the basis B and coordinatize the outputs relative to C,

$$\rho_C(T(1)) = \rho_C\left(\begin{bmatrix} 2 & 1\\ 1 & 1 \end{bmatrix}\right) = \rho_C\left(2\begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 2\\ 1\\ 1 \end{bmatrix}$$

$$\rho_C(T(1+x)) = \rho_C\left(\begin{bmatrix} 1 & 4\\ 4 & 1 \end{bmatrix}\right) = \rho_C\left(1\begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + 4\begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1\\ 4\\ 1 \end{bmatrix}$$

$$\rho_C\left(T\left(1+x+x^2\right)\right) = \rho_C\left(\begin{bmatrix} 2 & 3\\ 3 & 0 \end{bmatrix}\right) = \rho_C\left(2\begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + 3\begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 2\\ 3\\ 0 \end{bmatrix}$$

Applying Definition MR [1286] we have the matrix representation

$$M_{B,C}^T = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 4 & 3 \\ 1 & 1 & 0 \end{bmatrix}$$

To compute $T(3 + 5x - 2x^2)$ employ Theorem FTMR [1293],

$$T (3 + 5x - 2x^{2}) = \rho_{C}^{-1} (M_{B,C}^{T} \rho_{B} (3 + 5x - 2x^{2}))$$

= $\rho_{C}^{-1} (M_{B,C}^{T} \rho_{B} ((-2)(1) + 7(1 + x) + (-2)(1 + x + x^{2})))$
= $\rho_{C}^{-1} \left(\begin{bmatrix} 2 & 1 & 2 \\ 1 & 4 & 3 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 7 \\ -2 \end{bmatrix} \right)$

$$= \rho_C^{-1} \left(\begin{bmatrix} -1\\ 20\\ 5 \end{bmatrix} \right)$$

= (-1) $\begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + 20 \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 5 \begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix}$
= $\begin{bmatrix} -1 & 20\\ 20 & 5 \end{bmatrix}$

You can, of course, check your answer by evaluating $T(3 + 5x - 2x^2)$ directly. **C25** Contributed by Robert Beezer Statement [1332] Choose bases *B* and *C* for the matrix representation,

$$B = \{1, x, x^2, x^3\} \qquad C = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Input to T the vectors of the basis B and coordinatize the outputs relative to C,

$$\rho_C\left(T\left(1\right)\right) = \rho_C\left(\begin{bmatrix}-1 & 4\\1 & 1\end{bmatrix}\right) = \rho_C\left(\left(-1\right)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 4\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}-1\\4\\1\\1\end{bmatrix}$$

$$\rho_{C}(T(x)) = \rho_{C}\left(\begin{bmatrix}4 & -1\\5 & 0\end{bmatrix}\right) = \rho_{C}\left(4\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + (-1)\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 5\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 0\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}4\\-1\\5\\0\end{bmatrix}$$
$$\rho_{C}(T(x^{2})) = \rho_{C}\left(\begin{bmatrix}1 & 6\\-2 & 2\end{bmatrix}\right) = \rho_{C}\left(1\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 6\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + (-2)\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}1\\6\\-2\\2\end{bmatrix}$$
$$\rho_{C}(T(x^{3})) = \rho_{C}\left(\begin{bmatrix}2 & -1\\2 & 5\end{bmatrix}\right) = \rho_{C}\left(2\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + (-1)\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 5\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}2\\-1\\2\\5\end{bmatrix}$$

Applying Definition MR [1286] we have the matrix representation

$$M_{B,C}^{T} = \begin{bmatrix} -1 & 4 & 1 & 2\\ 4 & -1 & 6 & -1\\ 1 & 5 & -2 & 2\\ 1 & 0 & 2 & 5 \end{bmatrix}$$

Properties of this matrix representation will translate to properties of the linear transformation The matrix representation is nonsingular since it row-reduces to the identity matrix (Theorem NMRRI [176]) and therefore has a column space equal to \mathbb{C}^4 (Theorem CNMB [789]). The column space of the matrix representation is isomorphic to the range of the linear transformation (Theorem RCSI [1317]). So the range of T has dimension 4, equal to the dimension of the codomain M_{22} . By Theorem ROSLT [1234], T is surjective.

C30 Contributed by Robert Beezer Statement [1333] These subspaces will be easiest to construct by analyzing a matrix representation of S. Since we can use any matrix representation, we might as well use natural bases that allow us to construct the matrix representation quickly and easily,

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \qquad \qquad C = \left\{ 1, x, x^2 \right\}$$

then we can practically build the matrix representation on sight,

$$M_{B,C}^S = \begin{bmatrix} 1 & 2 & 5 & -4 \\ 3 & -1 & 8 & 2 \\ 1 & 1 & 4 & -2 \end{bmatrix}$$

The first step is to find bases for the null space and column space of the matrix representation. Row-reducing the matrix representation we find,

$$\begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So by Theorem BNS [334] and Theorem BCS [569], we have

$$\mathcal{N}(M_{B,C}^{S}) = \left\langle \left\{ \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\2\\0\\1 \end{bmatrix} \right\} \right\rangle \qquad \mathcal{C}(M_{B,C}^{S}) = \left\langle \left\{ \begin{bmatrix} 1\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\-1\\1 \end{bmatrix} \right\} \right\rangle$$

Now, the proofs of Theorem KNSI [1311] and Theorem RCSI [1317] tell us that we can apply ρ_B^{-1} and ρ_C^{-1} (respectively) to "un-coordinatize" and get bases for the kernel and range of the linear transformation S itself,

$$\mathcal{K}(S) = \left\langle \left\{ \begin{bmatrix} -3 & -1\\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2\\ 0 & 1 \end{bmatrix} \right\} \right\rangle \qquad \mathcal{R}(S) = \left\langle \left\{ 1 + 3x + x^2, 2 - x + x^2 \right\} \right\rangle$$

C40 Contributed by Robert Beezer Statement [1333]

The analysis of R will be easiest if we analyze a matrix representation of R. Since we can use any matrix representation, we might as well use natural bases that allow us to construct the matrix representation quickly and easily,

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \qquad \qquad C = \left\{ 1, x, x^2 \right\}$$

then we can practically build the matrix representation on sight,

$$M_{B,C}^{R} = \begin{bmatrix} 1 & -1 & 0\\ 2 & -3 & -2\\ 1 & -1 & 1 \end{bmatrix}$$

This matrix representation is invertible (it has a nonzero determinant of -1, Theorem SMZD [935], Theorem NI [542]) so Theorem IMR [1322] tells us that the linear transformation R is also invertible. To find a formula for R^{-1} we compute,

$$R^{-1} (a + bx + cx^{2}) = \rho_{B}^{-1} \left(M_{C,B}^{R^{-1}} \rho_{C} (a + bx + cx^{2}) \right)$$
 Theorem FTMR [1293]

$$= \rho_{B}^{-1} \left(\left(M_{B,C}^{R} \right)^{-1} \rho_{C} (a + bx + cx^{2}) \right)$$
 Theorem IMR [1322]

$$= \rho_{B}^{-1} \left(\left(M_{B,C}^{R} \right)^{-1} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right)$$
 Definition VR [1261]

$$= \rho_{B}^{-1} \left(\begin{bmatrix} 5 & -1 & -2 \\ 4 & -1 & -2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right)$$
 Definition MI [506]

$$= \rho_{B}^{-1} \left(\begin{bmatrix} 5a - b - 2c \\ 4a - b - 2c \\ -a + c \end{bmatrix} \right)$$
 Definition MVP [461]

$$= \begin{bmatrix} 5a - b - 2c & 4a - b - 2c \\ 4a - b - 2c & -a + c \end{bmatrix}$$
 Definition VR [1261]

C41 Contributed by Robert Beezer Statement [1333] First, build a matrix representation of S (Definition MR [1286]). We are free to choose whatever bases we wish, so we should choose ones that are easy to work with, such as

$$B = \{1, x\}$$
$$C = \{\begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \end{bmatrix}\}$$

The resulting matrix representation is then

$$M_{B,C}^T = \begin{bmatrix} 3 & 1\\ 2 & 1 \end{bmatrix}$$

this matrix is invertible, since it has a nonzero determinant, so by Theorem IMR [1322] the linear transformation S is invertible. We can use the matrix inverse and Theorem IMR [1322] to find a formula for the inverse linear transformation,

$$S^{-1}\left(\begin{bmatrix}a & b\end{bmatrix}\right) = \rho_B^{-1}\left(M_{C,B}^{S^{-1}}\rho_C\left(\begin{bmatrix}a & b\end{bmatrix}\right)\right) \qquad \text{Theorem FTMR [1293]}$$
$$= \rho_B^{-1}\left(\left(M_{B,C}^S\right)^{-1}\rho_C\left(\begin{bmatrix}a & b\end{bmatrix}\right)\right) \qquad \text{Theorem IMR [1322]}$$
$$= \rho_B^{-1}\left(\left(M_{B,C}^S\right)^{-1}\begin{bmatrix}a\\b\end{bmatrix}\right) \qquad \text{Definition VR [1261]}$$
$$= \rho_B^{-1}\left(\left(\begin{bmatrix}3 & 1\\2 & 1\end{bmatrix}\right)^{-1}\begin{bmatrix}a\\b\end{bmatrix}\right)$$
$$= \rho_B^{-1}\left(\left(\begin{bmatrix}1 & -1\\-2 & 3\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}\right)\right) \qquad \text{Definition MI [506]}$$

$$= \rho_B^{-1} \left(\begin{bmatrix} a - b \\ -2a + 3b \end{bmatrix} \right)$$
 Definition MVP [461]
$$= (a - b) + (-2a + 3b)x$$
 Definition VR [1261]

C42 Contributed by Robert Beezer Statement [1333] Choose bases B and C for M_{12} and M_{21} (respectively),

$$B = \left\{ \begin{bmatrix} 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \end{bmatrix} \right\} C = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

The resulting matrix representation is

$$M_{B,C}^R = \begin{bmatrix} 1 & 3\\ 4 & 11 \end{bmatrix}$$

This matrix is invertible (its determinant is nonzero, Theorem SMZD [935]), so by Theorem IMR [1322], we can compute the matrix representation of R^{-1} with a matrix inverse (Theorem TTMI [509]),

$$M_{C,B}^{R^{-1}} = \begin{bmatrix} 1 & 3\\ 4 & 11 \end{bmatrix}^{-1} = \begin{bmatrix} -11 & 3\\ 4 & -1 \end{bmatrix}$$

To obtain a general formula for R^{-1} , use Theorem FTMR [1293],

$$R^{-1}\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \rho_B^{-1}\left(M_{C,B}^{R^{-1}}\rho_C\left(\begin{bmatrix}x\\y\end{bmatrix}\right)\right)$$

$$= \rho_B^{-1} \left(\begin{bmatrix} -11 & 3\\ 4 & -1 \end{bmatrix} \begin{bmatrix} x\\ y \end{bmatrix} \right)$$
$$= \rho_B^{-1} \left(\begin{bmatrix} -11x + 3y\\ 4x - y \end{bmatrix} \right)$$
$$= \begin{bmatrix} -11x + 3y & 4x - y \end{bmatrix}$$

C50 Contributed by Robert Beezer Statement [1334]

As usual, build any matrix representation of L, most likely using a "nice" bases, such as

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$
$$C = \left\{ 1, x, x^2 \right\}$$

Then the matrix representation (Definition MR [1286]) is,

$$M_{B,C}^{L} = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 3 & 0 & 1 & -2 \\ -1 & 1 & 3 & 3 \end{bmatrix}$$

Theorem RCSI [1317] tells us that we can compute the column space of the matrix representation, then use the isomorphism ρ_C^{-1} to convert the column space of the matrix representation into the range of the linear transformation. So we first analyze the matrix representation,

$$\begin{bmatrix} 1 & 2 & 4 & 1 \\ 3 & 0 & 1 & -2 \\ -1 & 1 & 3 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

With three nonzero rows in the reduced row-echelon form of the matrix, we know the column space has dimension 3. Since P_2 has dimension 3 (Theorem DP [824]), the range must be all of P_2 . So any basis of P_2 would suffice as a basis for the range. For instance, C itself would be a correct answer.

A more laborious approach would be to use Theorem BCS [569] and choose the first three columns of the matrix representation as a basis for the range of the matrix representation. These could then be "un-coordinatized" with ρ_C^{-1} to yield a ("not nice") basis for P_2 .

C52 Contributed by Robert Beezer Statement [1334] Choose bases B and C for the matrix representation,

$$B = \{1, x, x^2\} \qquad C = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Input to T the vectors of the basis B and coordinatize the outputs relative to C,

$$\rho_{C}(T(1)) = \rho_{C}\left(\begin{bmatrix}1 & 2\\ -1 & 3\end{bmatrix}\right) = \rho_{C}\left(1\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + (-1)\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} + 3\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right) = \begin{bmatrix}1\\ 2\\ -1\\ 3\end{bmatrix}$$
$$\rho_{C}(T(x)) = \rho_{C}\left(\begin{bmatrix}2 & 2\\ 1 & 2\end{bmatrix}\right) = \rho_{C}\left(2\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right) = \begin{bmatrix}2\\ 2\\ 1\\ 2\end{bmatrix}$$

$$\rho_C\left(T\left(x^2\right)\right) = \rho_C\left(\begin{bmatrix}-2 & 0\\-4 & 2\end{bmatrix}\right) = \rho_C\left((-2)\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 0\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + (-4)\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}-2\\0\\-4\\2\end{bmatrix}$$

Applying Definition MR [1286] we have the matrix representation

$$M_{B,C}^{T} = \begin{bmatrix} 1 & 2 & -2 \\ 2 & 2 & 0 \\ -1 & 1 & -4 \\ 3 & 2 & 2 \end{bmatrix}$$

The null space of the matrix representation is isomorphic (via ρ_B) to the kernel of the linear transformation (Theorem KNSI [1311]). So we compute the null space of the matrix representation by first row-reducing the matrix to,

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Employing Theorem BNS [334] we have

$$\mathcal{N}(M_{B,C}^{T}) = \left\langle \left\{ \begin{bmatrix} -2\\2\\1 \end{bmatrix} \right\} \right\rangle$$

We only need to uncoordinatize this one basis vector to get a basis for $\mathcal{K}(T)$,

$$\mathcal{K}(T) = \left\langle \left\{ \rho_B^{-1} \left(\begin{bmatrix} -2\\2\\1 \end{bmatrix} \right) \right\} \right\rangle = \left\langle \left\{ -2 + 2x + x^2 \right\} \right\rangle$$

M20 Contributed by Robert Beezer Statement [1335] Build a matrix representation (Definition MR [1286]) with the set

$$B = \left\{1, x, x^2, \dots, x^n\right\}$$

employed as a basis of both the domain and codomain. Then

$$\rho_B(D(1)) = \rho_B(0) = \begin{bmatrix} 0\\0\\0\\\vdots\\0\\0 \end{bmatrix} \qquad \qquad \rho_B(D(x)) = \rho_B(1) = \begin{bmatrix} 1\\0\\0\\\vdots\\0\\0 \end{bmatrix}$$

$$\rho_B(D(x^3)) = \rho_B(3x^2) = \begin{bmatrix} 0\\0\\3\\\vdots\\0\\0 \end{bmatrix}$$

$$\rho_B\left(D\left(x^n\right)\right) = \rho_B\left(nx^{n-1}\right) = \begin{bmatrix} 0\\0\\0\\\vdots\\n\\0\end{bmatrix}$$

and the resulting matrix representation is

$$M_{B,B}^{D} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 2 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 3 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & n \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

This $(n + 1) \times (n + 1)$ matrix is very close to being in reduced row-echelon form. Multiply row *i* by $\frac{1}{i}$, for $1 \le i \le n$, to convert it to reduced row-echelon form. From this we can see that matrix representation $M_{B,B}^D$ has rank *n* and nullity 1. Applying Theorem RCSI [1317] and Theorem KNSI [1311] tells us that the linear transformation *D* will have the same values for the rank and nullity, as well. **T20** Contributed by Robert Beezer Statement [1336] Given the nonsingular $n \times n$ matrix A, create the linear transformation $T: \mathbb{C}^n \to \mathbb{C}^n$ defined by $T(\mathbf{x}) = A\mathbf{x}$. Then

A nonsingular \Leftrightarrow A invertibleTheorem NI [542] \Leftrightarrow T invertibleTheorem IMILT [1327] \Leftrightarrow T injective and surjectiveTheorem ILTIS [1221] \Leftrightarrow C linearly independent, andTheorem ILTB [1149]C spans \mathbb{C}^n Theorem SLTB [1192] \Leftrightarrow C basis for \mathbb{C}^n Definition B [777]

T80 Contributed by Robert Beezer Statement [1336] Suppose that $B = {\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_m}, C = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n}$ and $D = {\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_p}$ For convenience, set $M = M_{B,C}^T$, $m_{ij} = [M]_{ij}, 1 \le i \le n, 1 \le j \le m$, and similarly, set $N = M_{C,D}^S, n_{ij} = [N]_{ij}, 1 \le i \le p, 1 \le j \le n$. We want to learn about the matrix representation of $S \circ T \colon V \to W$ relative to B and D. We will examine a single (generic) entry of this representation.

$$\begin{bmatrix} M_{B,D}^{S \circ T} \end{bmatrix}_{ij} = \left[\rho_D \left((S \circ T) \left(\mathbf{u}_j \right) \right) \right]_i \qquad \text{Definition MR [1286]} \\ = \left[\rho_D \left(S \left(T \left(\mathbf{u}_j \right) \right) \right) \right]_i \qquad \text{Definition LTC [111]} \end{aligned}$$

Version 2.99

$$= \left[\rho_D \left(S \left(\sum_{k=1}^n m_{kj} \mathbf{v}_k \right) \right) \right]_i$$

$$= \left[\rho_D \left(\sum_{k=1}^n m_{kj} S \left(\mathbf{v}_k \right) \right) \right]_i$$

$$= \left[\rho_D \left(\sum_{k=1}^n m_{kj} \sum_{\ell=1}^p n_{\ell k} \mathbf{w}_\ell \right) \right]_i$$

$$= \left[\rho_D \left(\sum_{k=1}^n \sum_{\ell=1}^p m_{kj} n_{\ell k} \mathbf{w}_\ell \right) \right]_i$$

$$= \left[\rho_D \left(\sum_{\ell=1}^p \sum_{k=1}^n m_{kj} n_{\ell k} \mathbf{w}_\ell \right) \right]_i$$

$$= \sum_{k=1}^n m_{kj} n_{ik}$$

$$= \sum_{k=1}^n n_{ik} m_{kj}$$

Definition MR [1286]

Definition MR [1286]

Property DVA [660]

Property C [660]

Property DSA [660]

Definition VR [1261]

Property CMCN [1597]

$$= \sum_{k=1}^{n} \left[M_{C,D}^{S} \right]_{ik} \left[M_{B,C}^{T} \right]_{kj}$$
 Property CMCN [1597]

This formula for the entry of a matrix should remind you of Theorem EMP [471]. However, while the theorem presumed we knew how to multiply matrices, the solution before us never uses any understanding of matrix products. It uses the definitions of vector and matrix representations, properties of linear transformations and vector spaces. So if we began a course by first discussing vector space, and then linear transformations between vector spaces, we could carry matrix representations into a *motivation* for a definition of matrix multiplication that is grounded in function composition. That is worth saying again — a definition of matrix representations of linear transformations *results* in a matrix product being the representation of a composition of linear transformations.

This exercise is meant to explain why many authors take the formula in Theorem EMP [471] as their *definition* of matrix multiplication, and why it is a natural choice when the proper motivation is in place. If we first defined matrix multiplication in the style of Theorem EMP [471], then the above argument, followed by a simple application of the definition of matrix equality (Definition ME [427]), would yield Theorem MRCLT [1302].

Section CB Change of Basis

We have seen in Section MR [1286] that a linear transformation can be represented by a matrix, once we pick bases for the domain and codomain. How does the matrix representation change if we choose different bases? Which bases lead to especially nice representations? From the infinite possibilities, what is the best possible representation? This section will begin to answer these questions. But first we need to define eigenvalues for linear transformations and the change-of-basis matrix.

Subsection EELT Eigenvalues and Eigenvectors of Linear Transformations

We now define the notion of an eigenvalue and eigenvector of a linear transformation. It should not be too surprising, especially if you remind yourself of the close relationship between matrices and linear transformations.

Definition EELT

Eigenvalue and Eigenvector of a Linear Transformation

Suppose that $T: V \to V$ is a linear transformation. Then a nonzero vector $\mathbf{v} \in V$ is an **eigenvector** of T for the **eigenvalue** λ if $T(\mathbf{v}) = \lambda \mathbf{v}$.

We will see shortly the best method for computing the eigenvalues and eigenvectors of a linear transformation, but for now, here are some examples to verify that such things really do exist.

Example ELTBM

Eigenvectors of linear transformation between matrices

Consider the linear transformation $T: M_{22} \to M_{22}$ defined by

$$T\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right) = \begin{bmatrix}-17a + 11b + 8c - 11d & -57a + 35b + 24c - 33d\\-14a + 10b + 6c - 10d & -41a + 25b + 16c - 23d\end{bmatrix}$$

and the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \qquad \qquad \mathbf{x}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \qquad \qquad \mathbf{x}_3 = \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix} \qquad \qquad \mathbf{x}_4 = \begin{bmatrix} 2 & 6 \\ 1 & 4 \end{bmatrix}$$

Then compute

$$T(\mathbf{x}_1) = T\left(\begin{bmatrix} 0 & 1\\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 0 & 2\\ 0 & 2 \end{bmatrix} = 2\mathbf{x}_1$$
$$T(\mathbf{x}_2) = T\left(\begin{bmatrix} 1 & 1\\ 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} 2 & 2\\ 2 & 0 \end{bmatrix} = 2\mathbf{x}_2$$

$$T (\mathbf{x}_3) = T \left(\begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix} \right) = \begin{bmatrix} -1 & -3 \\ -2 & -3 \end{bmatrix} = (-1)\mathbf{x}_3$$
$$T (\mathbf{x}_4) = T \left(\begin{bmatrix} 2 & 6 \\ 1 & 4 \end{bmatrix} \right) = \begin{bmatrix} -4 & -12 \\ -2 & -8 \end{bmatrix} = (-2)\mathbf{x}_4$$

So \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 , \mathbf{x}_4 are eigenvectors of T with eigenvalues (respectively) $\lambda_1 = 2$, $\lambda_2 = 2$, $\lambda_3 = -1$, $\lambda_4 = -2$.

Here's another.

Example ELTBP

Eigenvectors of linear transformation between polynomials Consider the linear transformation $R: P_2 \to P_2$ defined by

$$R(a + bx + cx^{2}) = (15a + 8b - 4c) + (-12a - 6b + 3c)x + (24a + 14b - 7c)x^{2}$$

and the vectors

$$\mathbf{w}_1 = 1 - x + x^2$$
 $\mathbf{w}_2 = x + 2x^2$ $\mathbf{w}_3 = 1 + 4x^2$

Then compute

$$R(\mathbf{w}_{1}) = R(1 - x + x^{2}) = 3 - 3x + 3x^{2} = 3\mathbf{w}_{1}$$

$$R(\mathbf{w}_{2}) = R(x + 2x^{2}) = 0 + 0x + 0x^{2} = 0\mathbf{w}_{2}$$

$$R(\mathbf{w}_3) = R(1+4x^2) = -1 - 4x^2 = (-1)\mathbf{w}_3$$

So \mathbf{w}_1 , \mathbf{w}_2 , \mathbf{w}_3 are eigenvectors of R with eigenvalues (respectively) $\lambda_1 = 3$, $\lambda_2 = 0$, $\lambda_3 = -1$. Notice how the eigenvalue $\lambda_2 = 0$ indicates that the eigenvector \mathbf{w}_2 is a non-trivial element of the kernel of R, and therefore R is not injective (Exercise CB.T15 [1405]).

Of course, these examples are meant only to illustrate the definition of eigenvectors and eigenvalues for linear transformations, and therefore beg the question, "How would I *find* eigenvectors?" We'll have an answer before we finish this section. We need one more construction first.

Subsection CBM Change-of-Basis Matrix

Given a vector space, we know we can usually find many different bases for the vector space, some nice, some nasty. If we choose a single vector from this vector space, we can build many different representations of the vector by constructing the representations relative to different bases. How are these different representations related to each other? A change-of-basis matrix answers this question.

Definition CBM

Change-of-Basis Matrix

Suppose that V is a vector space, and $I_V: V \to V$ is the identity linear transformation on V. Let $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n}$ and C be two bases of V. Then the **change-of-basis matrix** from B to C is the matrix representation of I_V relative to B and C,

$$C_{B,C} = M_{B,C}^{I_V}$$

= $[\rho_C (I_V (\mathbf{v}_1)) | \rho_C (I_V (\mathbf{v}_2)) | \rho_C (I_V (\mathbf{v}_3)) | \dots | \rho_C (I_V (\mathbf{v}_n))]$
= $[\rho_C (\mathbf{v}_1) | \rho_C (\mathbf{v}_2) | \rho_C (\mathbf{v}_3) | \dots | \rho_C (\mathbf{v}_n)]$

Notice that this definition is primarily about a single vector space (V) and two bases of V(B, C). The linear transformation (I_V) is necessary but not critical. As you might expect, this matrix has something to do with changing bases. Here is the theorem that gives the matrix its name (not the other way around).

Theorem CB Change-of-Basis

Suppose that \mathbf{v} is a vector in the vector space V and B and C are bases of V. Then

$$\rho_{C}\left(\mathbf{v}\right) = C_{B,C}\rho_{B}\left(\mathbf{v}\right)$$

Proof

$$\rho_{C}(\mathbf{v}) = \rho_{C}(I_{V}(\mathbf{v}))$$
Definition IDLT [1213]

 \triangle

$$= M_{B,C}^{I_V} \rho_B (\mathbf{v})$$
 Theorem FTMR [1293]
$$= C_{B,C} \rho_B (\mathbf{v})$$
 Definition CBM [1359]

So the change-of-basis matrix can be used with matrix multiplication to convert a vector representation of a vector (**v**) relative to one basis ($\rho_B(\mathbf{v})$) to a representation of the same vector relative to a second basis ($\rho_C(\mathbf{v})$).

Theorem ICBM

Inverse of Change-of-Basis Matrix

Suppose that V is a vector space, and B and C are bases of V. Then the change-of-basis matrix $C_{B,C}$ is nonsingular and

$$C_{B,C}^{-1} = C_{C,B}$$

Proof The linear transformation $I_V: V \to V$ is invertible, and its inverse is itself, I_V (check this!). So by Theorem IMR [1322], the matrix $M_{B,C}^{I_V} = C_{B,C}$ is invertible. Theorem NI [542] says an invertible matrix is nonsingular.

Then

$$C_{B,C}^{-1} = \left(M_{B,C}^{I_V}\right)^{-1} \\ = M_{C,B}^{I_V^{-1}} \\ = M_{C,B}^{I_V}$$

Definition CBM [1359] Theorem IMR [1322] Definition IDLT [1213] $= C_{C,B}$ Definition CBM [1359]

Example CBP Change of basis with polynomials

The vector space P_4 (Example VSP [665]) has two nice bases (Example BP [780]),

$$B = \left\{1, x, x^2, x^3, x^4\right\} \quad C = \left\{1, 1 + x, 1 + x + x^2, 1 + x + x^2 + x^3, 1 + x + x^2 + x^3 + x^4\right\}$$

To build the change-of-basis matrix between B and C, we must first build a vector representation of each vector in B relative to C,

$$\rho_{C}(1) = \rho_{C}((1)(1)) = \begin{bmatrix} 1\\0\\0\\0\\0 \end{bmatrix}$$

$$\rho_{C}(x) = \rho_{C}((-1)(1) + (1)(1+x)) = \begin{bmatrix} -1\\1\\0\\0\\0\\0 \end{bmatrix}$$

_

$$\rho_C \left(x^2\right) = \rho_C \left(\left(-1\right)\left(1+x\right) + \left(1\right)\left(1+x+x^2\right)\right) = \begin{bmatrix} 0\\-1\\1\\0\\0 \end{bmatrix}$$

$$\rho_C \left(x^3\right) = \rho_C \left(\left(-1\right)\left(1+x+x^2\right) + \left(1\right)\left(1+x+x^2+x^3\right)\right) = \begin{bmatrix} 0\\0\\-1\\1\\0 \end{bmatrix}$$

$$\rho_C \left(x^4\right) = \rho_C \left(\left(-1\right)\left(1+x+x^2+x^3\right) + \left(1\right)\left(1+x+x^2+x^3+x^4\right)\right) = \begin{bmatrix} 0\\0\\-1\\1\\1 \end{bmatrix}$$

Then we package up these vectors as the columns of a matrix,

$$C_{B,C} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Now, to illustrate Theorem CB [1359], consider the vector $\mathbf{u} = 5 - 3x + 2x^2 + 8x^3 - 3x^4$. We can build the representation of \mathbf{u} relative to B easily,

$$\rho_B \left(\mathbf{u} \right) = \rho_B \left(5 - 3x + 2x^2 + 8x^3 - 3x^4 \right) = \begin{bmatrix} 5 \\ -3 \\ 2 \\ 8 \\ -3 \end{bmatrix}$$

Applying Theorem CB [1359], we obtain a second representation of \mathbf{u} , but now relative to C,

$$\rho_{C} (\mathbf{u}) = C_{B,C} \rho_{B} (\mathbf{u}) \qquad \text{Theorem CB [1359]} \\
= \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \\ 2 \\ 8 \\ -3 \end{bmatrix} \\
= \begin{bmatrix} 8 \\ -5 \\ -6 \\ 11 \\ -3 \end{bmatrix} \qquad \text{Definition MVP [461]}$$

We can check our work by unraveling this second representation,

$$\begin{aligned} \mathbf{u} &= \rho_C^{-1} \left(\rho_C \left(\mathbf{u} \right) \right) & \text{Definition IVLT [1214]} \\ &= \rho_C^{-1} \left(\begin{bmatrix} 8 \\ -5 \\ -6 \\ 11 \\ -3 \end{bmatrix} \right) \\ &= 8(1) + (-5)(1+x) + (-6)(1+x+x^2) \\ &+ (11)(1+x+x^2+x^3) + (-3)(1+x+x^2+x^3+x^4) \\ &= 5 - 3x + 2x^2 + 8x^3 - 3x^4 \end{aligned}$$

The change-of-basis matrix from C to B is actually easier to build. Grab each vector in the basis C and form its representation relative to B

$$\rho_B(1) = \rho_B((1)1) = \begin{bmatrix} 1\\ 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix}$$

$$\rho_B (1+x) = \rho_B ((1)1+(1)x) = \begin{bmatrix} 1\\1\\0\\0\\0\end{bmatrix}$$

$$\rho_B (1+x+x^2) = \rho_B ((1)1+(1)x+(1)x^2) = \begin{bmatrix} 1\\1\\0\\0\end{bmatrix}$$

$$\rho_B (1+x+x^2+x^3) = \rho_B ((1)1+(1)x+(1)x^2+(1)x^3) = \begin{bmatrix} 1\\1\\1\\0\\0\end{bmatrix}$$

$$\rho_B (1+x+x^2+x^3+x^4) = \rho_B ((1)1+(1)x+(1)x^2+(1)x^3+(1)x^4) = \begin{bmatrix} 1\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

Then we package up these vectors as the columns of a matrix,

$$C_{C,B} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We formed two representations of the vector \mathbf{u} above, so we can again provide a check on our computations by converting from the representation of \mathbf{u} relative to C to the representation of \mathbf{u} relative to B,

$$\rho_B (\mathbf{u}) = C_{C,B} \rho_C (\mathbf{u}) \qquad \text{Theorem CB [1359]} \\
= \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8 \\ -5 \\ -6 \\ 11 \\ -3 \end{bmatrix} \\
= \begin{bmatrix} 5 \\ -3 \\ 2 \\ 8 \\ -3 \end{bmatrix} \qquad \text{Definition MVP [461]}$$

One more computation that is either a check on our work, or an illustration of a theorem. The two change-of-basis matrices, $C_{B,C}$ and $C_{C,B}$, should be inverses of each other, according to Theorem ICBM [1360]. Here we go,

$$C_{B,C}C_{C,B} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The computations of the previous example are not meant to present any labor-saving devices, but instead are meant to illustrate the *utility* of the change-of-basis matrix. However, you might have noticed that $C_{C,B}$ was easier to compute than $C_{B,C}$. If you needed $C_{B,C}$, then you could first compute $C_{C,B}$ and then compute its inverse, which by Theorem ICBM [1360], would equal $C_{B,C}$.

Here's another illustrative example. We have been concentrating on working with abstract vector spaces, but all of our theorems and techniques apply just as well to \mathbb{C}^m , the vector space of column vectors. We only need to use more complicated bases than the standard unit vectors (Theorem SUVB [779]) to make things interesting.

Example CBCV Change of basis with column vectors \boxtimes

For the vector space \mathbb{C}^4 we have the two bases,

$$B = \left\{ \begin{bmatrix} 1\\-2\\1\\-2 \end{bmatrix}, \begin{bmatrix} -1\\3\\1\\1 \end{bmatrix}, \begin{bmatrix} 2\\-3\\3\\-4 \end{bmatrix}, \begin{bmatrix} -1\\3\\3\\0 \end{bmatrix} \right\} \qquad C = \left\{ \begin{bmatrix} 1\\-6\\-4\\-1 \end{bmatrix}, \begin{bmatrix} -4\\8\\-5\\8 \end{bmatrix}, \begin{bmatrix} -5\\13\\-2\\9 \end{bmatrix}, \begin{bmatrix} 3\\-7\\3\\-6 \end{bmatrix} \right\}$$

The change-of-basis matrix from B to C requires writing each vector of B as a linear combination the vectors in C,

$$\rho_{C} \left(\begin{bmatrix} 1\\-2\\1\\-2 \end{bmatrix} \right) = \rho_{C} \left((1) \begin{bmatrix} 1\\-6\\-4\\-1 \end{bmatrix} + (-2) \begin{bmatrix} -4\\8\\-5\\8 \end{bmatrix} + (1) \begin{bmatrix} -5\\13\\-2\\9 \end{bmatrix} + (-1) \begin{bmatrix} 3\\-7\\3\\-6 \end{bmatrix} \right) = \begin{bmatrix} 1\\-2\\1\\-7\\1\\-1 \end{bmatrix}$$

$$\rho_{C} \left(\begin{bmatrix} -1\\3\\1\\1 \end{bmatrix} \right) = \rho_{C} \left((2) \begin{bmatrix} 1\\-6\\-4\\-1 \end{bmatrix} + (-3) \begin{bmatrix} -4\\8\\-5\\8 \end{bmatrix} + (3) \begin{bmatrix} -5\\13\\-2\\9 \end{bmatrix} + (0) \begin{bmatrix} 3\\-7\\3\\-6 \end{bmatrix} \right) = \begin{bmatrix} 2\\-3\\3\\0 \end{bmatrix}$$

$$\rho_{C} \left(\begin{bmatrix} 2\\-3\\3\\-6 \end{bmatrix} \right) = \rho_{C} \left((1) \begin{bmatrix} 1\\-6\\-4\\-1 \end{bmatrix} + (-3) \begin{bmatrix} -4\\8\\-5\\8 \end{bmatrix} + (1) \begin{bmatrix} -5\\13\\-2\\9 \end{bmatrix} + (-2) \begin{bmatrix} 3\\-7\\3\\-6 \end{bmatrix} \right) = \begin{bmatrix} 1\\-3\\1\\-2\\1 \end{bmatrix}$$

$$\rho_C \left(\begin{bmatrix} -1\\3\\3\\0 \end{bmatrix} \right) = \rho_C \left((2) \begin{bmatrix} 1\\-6\\-4\\-1 \end{bmatrix} + (-2) \begin{bmatrix} -4\\8\\-5\\8 \end{bmatrix} + (4) \begin{bmatrix} -5\\13\\-2\\9 \end{bmatrix} + (3) \begin{bmatrix} 3\\-7\\3\\-6 \end{bmatrix} \right) = \begin{bmatrix} 2\\-2\\4\\3 \end{bmatrix}$$

Then we package these vectors up as the change-of-basis matrix,

$$C_{B,C} = \begin{bmatrix} 1 & 2 & 1 & 2 \\ -2 & -3 & -3 & -2 \\ 1 & 3 & 1 & 4 \\ -1 & 0 & -2 & 3 \end{bmatrix}$$

Now consider a single (arbitrary) vector $\mathbf{y} = \begin{bmatrix} 2 \\ 6 \\ -3 \\ 4 \end{bmatrix}$. First, build the vector representation

of y relative to B. This will require writing y as a linear combination of the vectors in B,

$$\rho_B\left(\mathbf{y}\right) = \rho_B\left(\begin{bmatrix}2\\6\\-3\\4\end{bmatrix}\right)$$

$$= \rho_B \left((-21) \begin{bmatrix} 1\\-2\\1\\-2 \end{bmatrix} + (6) \begin{bmatrix} -1\\3\\1\\1 \end{bmatrix} + (11) \begin{bmatrix} 2\\-3\\-3\\-4 \end{bmatrix} + (-7) \begin{bmatrix} -1\\3\\-3\\0 \end{bmatrix} \right) = \begin{bmatrix} -21\\6\\11\\-7 \end{bmatrix}$$

Now, applying Theorem CB [1359] we can convert the representation of \mathbf{y} relative to B into a representation relative to C,

$$\rho_{C} (\mathbf{y}) = C_{B,C} \rho_{B} (\mathbf{y})$$
Theorem CB [1359]
$$= \begin{bmatrix} 1 & 2 & 1 & 2 \\ -2 & -3 & -3 & -2 \\ 1 & 3 & 1 & 4 \\ -1 & 0 & -2 & 3 \end{bmatrix} \begin{bmatrix} -21 \\ 6 \\ 11 \\ -7 \end{bmatrix}$$

$$= \begin{bmatrix} -12 \\ 5 \\ -20 \\ -22 \end{bmatrix}$$
Definition MVP [461]

We could continue further with this example, perhaps by computing the representation of \mathbf{y} relative to the basis C directly as a check on our work (Exercise CB.C20 [1404]). Or we could choose another vector to play the role of \mathbf{y} and compute two different representations of this vector relative to the two bases B and C.

 \boxtimes

Subsection MRS Matrix Representations and Similarity

Here is the main theorem of this section. It looks a bit involved at first glance, but the proof should make you realize it is not all that complicated. In any event, we are more interested in a special case.

Theorem MRCB

Matrix Representation and Change of Basis

Suppose that $T: U \to V$ is a linear transformation, B and C are bases for U, and D and E are bases for V. Then

$$M_{B,D}^T = C_{E,D} M_{C,E}^T C_{B,C}$$

Proof

$$C_{E,D}M_{C,E}^{T}C_{B,C} = M_{E,D}^{I_{V}}M_{C,E}^{T}M_{B,C}^{I_{U}}$$
$$= M_{E,D}^{I_{V}}M_{B,E}^{T \circ I_{U}}$$
$$= M_{E,D}^{I_{V}}M_{B,E}^{T}$$
$$= M_{B,D}^{I_{V} \circ T}$$
$$= M_{B,D}^{T}$$

Definition CBM [1359] Theorem MRCLT [1302] Definition IDLT [1213] Theorem MRCLT [1302] Definition IDLT [1213] \square

We will be most interested in a special case of this theorem (Theorem SCB [1377]), but here's an example that illustrates the full generality of Theorem MRCB [1371].

Example MRCM

Matrix representations and change-of-basis matrices

Begin with two vector spaces, S_2 , the subspace of M_{22} containing all 2×2 symmetric matrices, and P_3 (Example VSP [665]), the vector space of all polynomials of degree 3 or less. Then define the linear transformation $Q: S_2 \to P_3$ by

$$Q\left(\begin{bmatrix}a & b\\ b & c\end{bmatrix}\right) = (5a - 2b + 6c) + (3a - b + 2c)x + (a + 3b - c)x^2 + (-4a + 2b + c)x^3$$

Here are two bases for each vector space, one nice, one nasty. First for S_2 ,

$$B = \left\{ \begin{bmatrix} 5 & -3 \\ -3 & -2 \end{bmatrix}, \begin{bmatrix} 2 & -3 \\ -3 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \right\} \qquad C = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

and then for P_3 ,

$$D = \left\{2 + x - 2x^2 + 3x^3, \ -1 - 2x^2 + 3x^3, \ -3 - x + x^3, \ -x^2 + x^3\right\} \quad E = \left\{1, \ x, \ x^2, \ x^3\right\}$$

We'll begin with a matrix representation of Q relative to C and E. We first find vector

representations of the elements of C relative to E,

$$\rho_E \left(Q \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \right) = \rho_E \left(5 + 3x + x^2 - 4x^3 \right) = \begin{bmatrix} 5 \\ 3 \\ 1 \\ -4 \end{bmatrix}$$
$$\rho_E \left(Q \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \right) = \rho_E \left(-2 - x + 3x^2 + 2x^3 \right) = \begin{bmatrix} -2 \\ -1 \\ 3 \\ 2 \end{bmatrix}$$
$$\rho_E \left(Q \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) \right) = \rho_E \left(6 + 2x - x^2 + x^3 \right) = \begin{bmatrix} 6 \\ 2 \\ -1 \\ 1 \end{bmatrix}$$

 So

$$M^Q_{C,E} = \begin{bmatrix} 5 & -2 & 6 \\ 3 & -1 & 2 \\ 1 & 3 & -1 \\ -4 & 2 & 1 \end{bmatrix}$$

Now we construct two change-of-basis matrices. First, $C_{B,C}$ requires vector representations

of the elements of B, relative to C. Since C is a nice basis, this is straightforward,

$$\rho_{C} \left(\begin{bmatrix} 5 & -3 \\ -3 & -2 \end{bmatrix} \right) = \rho_{C} \left((5) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (-3) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + (-2) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 5 \\ -3 \\ -2 \end{bmatrix}$$
$$\rho_{C} \left(\begin{bmatrix} 2 & -3 \\ -3 & 0 \end{bmatrix} \right) = \rho_{C} \left((2) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (-3) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + (0) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix}$$
$$\rho_{C} \left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \right) = \rho_{C} \left((1) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + (2) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + (4) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

So

$$C_{B,C} = \begin{bmatrix} 5 & 2 & 1 \\ -3 & -3 & 2 \\ -2 & 0 & 4 \end{bmatrix}$$

The other change-of-basis matrix we'll compute is $C_{E,D}$. However, since E is a nice basis (and D is not) we'll turn it around and instead compute $C_{D,E}$ and apply Theorem ICBM [1360] to use an inverse to compute $C_{E,D}$.

$$\rho_E \left(2 + x - 2x^2 + 3x^3 \right) = \rho_E \left((2)1 + (1)x + (-2)x^2 + (3)x^3 \right) = \begin{bmatrix} 2\\1\\-2\\3 \end{bmatrix}$$

$$\rho_E \left(-1 - 2x^2 + 3x^3 \right) = \rho_E \left((-1)1 + (0)x + (-2)x^2 + (3)x^3 \right) = \begin{bmatrix} -1\\0\\-2\\3 \end{bmatrix}$$
$$\rho_E \left(-3 - x + x^3 \right) = \rho_E \left((-3)1 + (-1)x + (0)x^2 + (1)x^3 \right) = \begin{bmatrix} -3\\-1\\0\\1 \end{bmatrix}$$
$$\rho_E \left(-x^2 + x^3 \right) = \rho_E \left((0)1 + (0)x + (-1)x^2 + (1)x^3 \right) = \begin{bmatrix} 0\\0\\-1\\1 \end{bmatrix}$$

So, we can package these column vectors up as a matrix to obtain $C_{D,E}$ and then,

$$C_{E,D} = (C_{D,E})^{-1}$$
$$= \begin{bmatrix} 2 & -1 & -3 & 0 \\ 1 & 0 & -1 & 0 \\ -2 & -2 & 0 & -1 \\ 3 & 3 & 1 & 1 \end{bmatrix}^{-1}$$

Theorem ICBM [1360]

$$= \begin{bmatrix} 1 & -2 & 1 & 1 \\ -2 & 5 & -1 & -1 \\ 1 & -3 & 1 & 1 \\ 2 & -6 & -1 & 0 \end{bmatrix}$$

We are now in a position to apply Theorem MRCB [1371]. The matrix representation of Q relative to B and D can be obtained as follows,

$$\begin{split} M_{B,D}^{Q} &= C_{E,D} M_{C,E}^{Q} C_{B,C} & \text{Theorem MRCB [1371]} \\ &= \begin{bmatrix} 1 & -2 & 1 & 1 \\ -2 & 5 & -1 & -1 \\ 1 & -3 & 1 & 1 \\ 2 & -6 & -1 & 0 \end{bmatrix} \begin{bmatrix} 5 & -2 & 6 \\ 3 & -1 & 2 \\ 1 & 3 & -1 \\ -4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5 & 2 & 1 \\ -3 & -3 & 2 \\ -2 & 0 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -2 & 1 & 1 \\ -2 & 5 & -1 & -1 \\ 1 & -3 & 1 & 1 \\ 2 & -6 & -1 & 0 \end{bmatrix} \begin{bmatrix} 19 & 16 & 25 \\ 14 & 9 & 9 \\ -2 & -7 & 3 \\ -28 & -14 & 4 \end{bmatrix} \\ &= \begin{bmatrix} -39 & -23 & 14 \\ 62 & 34 & -12 \\ -53 & -32 & 5 \\ -44 & -15 & -7 \end{bmatrix} \end{split}$$

Now check our work by computing $M_{B,D}^Q$ directly (Exercise CB.C21 [1404]).

Here is a special case of the previous theorem, where we choose U and V to be the same

vector space, so the matrix representations and the change-of-basis matrices are all square of the same size.

Theorem SCB Similarity and Change of Basis

Suppose that $T: V \to V$ is a linear transformation and B and C are bases of V. Then

$$M_{B,B}^{T} = C_{B,C}^{-1} M_{C,C}^{T} C_{B,C}$$

Proof In the conclusion of Theorem MRCB [1371], replace D by B, and replace E by C,

$$M_{B,B}^{T} = C_{C,B} M_{C,C}^{T} C_{B,C}$$
Theorem MRCB [1371]
= $C_{B,C}^{-1} M_{C,C}^{T} C_{B,C}$ Theorem ICBM [1360]

This is the third surprise of this chapter. Theorem SCB [1377] considers the special case where a linear transformation has the same vector space for the domain and codomain (V). We build a matrix representation of T using the basis B simultaneously for both the domain and codomain $(M_{B,B}^T)$, and then we build a second matrix representation of T, now using the basis C for both the domain and codomain $(M_{C,C}^T)$. Then these two representations are related via a similarity transformation (Definition SIM [1029]) using a change-of-basis matrix $(C_{B,C})!$

Example MRBE Matrix representation with basis of eigenvectors

We return to the linear transformation $T: M_{22} \to M_{22}$ of Example ELTBM [1356] defined by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} -17a + 11b + 8c - 11d & -57a + 35b + 24c - 33d \\ -14a + 10b + 6c - 10d & -41a + 25b + 16c - 23d \end{bmatrix}$$

In Example ELTBM [1356] we showcased four eigenvectors of T. We will now put these four vectors in a set,

$$B = \{\mathbf{x}_1, \, \mathbf{x}_2, \, \mathbf{x}_3, \, \mathbf{x}_4\} = \left\{ \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 2 & 3 \end{bmatrix}, \begin{bmatrix} 2 & 6 \\ 1 & 4 \end{bmatrix} \right\}$$

Check that B is a basis of M_{22} by first establishing the linear independence of B and then employing Theorem G [852] to get the spanning property easily. Here is a second set of 2×2 matrices, which also forms a basis of M_{22} (Example BM [780]),

$$C = \{\mathbf{y}_1, \, \mathbf{y}_2, \, \mathbf{y}_3, \, \mathbf{y}_4\} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

We can build two matrix representations of T, one relative to B and one relative to C. Each is easy, but for wildly different reasons. In our computation of the matrix representation relative to B we borrow some of our work in Example ELTBM [1356]. Here are the representations, then the explanation.

$$\begin{aligned} \rho_B \left(T \left(\mathbf{x}_1 \right) \right) &= \rho_B \left(2\mathbf{x}_1 \right) = \rho_B \left(2\mathbf{x}_1 + 0\mathbf{x}_2 + 0\mathbf{x}_3 + 0\mathbf{x}_4 \right) = \begin{bmatrix} 2\\0\\0\\0\\0 \end{bmatrix} \\ \rho_B \left(T \left(\mathbf{x}_2 \right) \right) &= \rho_B \left(2\mathbf{x}_2 \right) = \rho_B \left(0\mathbf{x}_1 + 2\mathbf{x}_2 + 0\mathbf{x}_3 + 0\mathbf{x}_4 \right) = \begin{bmatrix} 0\\2\\0\\0\\0\\0\\0 \end{bmatrix} \\ \rho_B \left(T \left(\mathbf{x}_3 \right) \right) &= \rho_B \left((-1)\mathbf{x}_3 \right) = \rho_B \left(0\mathbf{x}_1 + 0\mathbf{x}_2 + (-1)\mathbf{x}_3 + 0\mathbf{x}_4 \right) = \begin{bmatrix} 0\\0\\-1\\0\\0\\-1\\0\\0\\-2 \end{bmatrix} \end{aligned}$$

So the resulting representation is

$$M_{B,B}^{T} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

Very pretty. Now for the matrix representation relative to C first compute,

$$\begin{split} \rho_C \left(T \left(\mathbf{y}_1 \right) \right) &= \rho_C \left(\begin{bmatrix} -17 & -57 \\ -14 & -41 \end{bmatrix} \right) \\ &= \rho_C \left(\left(-17 \right) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \left(-57 \right) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \left(-14 \right) \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \left(-41 \right) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} -17 \\ -57 \\ -14 \\ -41 \end{bmatrix} \\ \rho_C \left(T \left(\mathbf{y}_2 \right) \right) &= \rho_C \left(\begin{bmatrix} 11 & 35 \\ 10 & 25 \end{bmatrix} \right) \\ &= \rho_C \left(11 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 35 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 10 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 25 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 11 \\ 35 \\ 10 \\ 25 \end{bmatrix} \\ \rho_C \left(T \left(\mathbf{y}_3 \right) \right) &= \rho_C \left(\begin{bmatrix} 8 & 24 \\ 6 & 16 \end{bmatrix} \right) \end{split}$$

_

$$= \rho_C \left(8 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 24 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 6 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 16 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 8 \\ 24 \\ 6 \\ 16 \end{bmatrix}$$
$$\rho_C \left(T \left(\mathbf{y}_4 \right) \right) = \rho_C \left(\begin{bmatrix} -11 & -33 \\ -10 & -23 \end{bmatrix} \right)$$
$$= \rho_C \left(\left(-11 \right) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \left(-33 \right) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \left(-10 \right) \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \left(-23 \right) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} -11 \\ -33 \\ -10 \\ -23 \end{bmatrix}$$

So the resulting representation is

$$M_{C,C}^{T} = \begin{bmatrix} -17 & 11 & 8 & -11 \\ -57 & 35 & 24 & -33 \\ -14 & 10 & 6 & -10 \\ -41 & 25 & 16 & -23 \end{bmatrix}$$

Not quite as pretty. The purpose of this example is to illustrate Theorem SCB [1377]. This theorem says that the two matrix representations, $M_{B,B}^T$ and $M_{C,C}^T$, of the one linear transformation, T, are related by a similarity transformation using the change-of-basis matrix $C_{B,C}$. Lets compute this change-of-basis matrix. Notice that since C is such a nice basis,

this is fairly straightforward,

$$\rho_{C}(\mathbf{x}_{1}) = \rho_{C}\left(\begin{bmatrix}0 & 1\\0 & 1\end{bmatrix}\right) = \rho_{C}\left(0\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 0\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}0\\1\\0\\1\end{bmatrix}$$

$$\rho_{C}(\mathbf{x}_{2}) = \rho_{C}\left(\begin{bmatrix}1 & 1\\1 & 0\end{bmatrix}\right) = \rho_{C}\left(1\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 0\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}1\\1\\1\\0\end{bmatrix}$$

$$\rho_{C}(\mathbf{x}_{3}) = \rho_{C}\left(\begin{bmatrix}1 & 3\\2 & 3\end{bmatrix}\right) = \rho_{C}\left(1\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 3\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 2\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 3\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}1\\3\\2\\3\end{bmatrix}$$

$$\rho_{C}(\mathbf{x}_{4}) = \rho_{C}\left(\begin{bmatrix}2 & 6\\1 & 4\end{bmatrix}\right) = \rho_{C}\left(2\begin{bmatrix}1 & 0\\0 & 0\end{bmatrix} + 6\begin{bmatrix}0 & 1\\0 & 0\end{bmatrix} + 1\begin{bmatrix}0 & 0\\1 & 0\end{bmatrix} + 4\begin{bmatrix}0 & 0\\0 & 1\end{bmatrix}\right) = \begin{bmatrix}2\\6\\1\\4\end{bmatrix}$$

So we have,

$$C_{B,C} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 1 & 3 & 6 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 3 & 4 \end{bmatrix}$$

Now, according to Theorem SCB [1377] we can write,

$$M_{B,B}^{T} = C_{B,C}^{-1} M_{C,C}^{T} C_{B,C}$$

$$\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 1 & 3 & 6 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 3 & 4 \end{bmatrix}^{-1} \begin{bmatrix} -17 & 11 & 8 & -11 \\ -57 & 35 & 24 & -33 \\ -14 & 10 & 6 & -10 \\ -41 & 25 & 16 & -23 \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 1 & 3 & 6 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 3 & 4 \end{bmatrix}$$

This should look and feel exactly like the process for diagonalizing a matrix, as was described in Section SD [1029]. And it is. \boxtimes

We can now return to the question of computing an eigenvalue or eigenvector of a linear transformation. For a linear transformation of the form $T: V \to V$, we know that representations relative to different bases are similar matrices. We also know that similar matrices have equal characteristic polynomials by Theorem SMEE [1035]. We will now show that eigenvalues of a linear transformation T are precisely the eigenvalues of *any* matrix representation of T. Since the choice of a different matrix representation leads to a similar matrix, there will be no "new" eigenvalues obtained from this second representation. Similarly, the change-of-basis matrix can be used to show that eigenvectors obtained from one matrix representation will be precisely those obtained from any other representation. So we can determine the eigenvalues and eigenvectors of a linear transformation by forming one matrix representation, using *any* basis we please, and analyzing the matrix in the manner of Chapter E [944].

Theorem EER

Eigenvalues, Eigenvectors, Representations

Suppose that $T: V \to V$ is a linear transformation and B is a basis of V. Then $\mathbf{v} \in V$ is an eigenvector of T for the eigenvalue λ if and only if $\rho_B(\mathbf{v})$ is an eigenvector of $M_{B,B}^T$ for the eigenvalue λ .

Proof (\Rightarrow) Assume that $\mathbf{v} \in V$ is an eigenvector of T for the eigenvalue λ . Then

$M_{B,B}^{T}\rho_{B}\left(\mathbf{v}\right) = \rho_{B}\left(T\left(\mathbf{v}\right)\right)$	Theorem FTMR [1293]
$= ho_{B}\left(\lambda\mathbf{v} ight)$	Definition EELT [1356]
$=\lambda ho_{B}\left(\mathbf{v} ight)$	Theorem VRLT [1262]

which by Definition EEM [945] says that $\rho_B(\mathbf{v})$ is an eigenvector of the matrix $M_{B,B}^T$ for the eigenvalue λ .

 (\Leftarrow) Assume that $\rho_B(\mathbf{v})$ is an eigenvector of $M_{B,B}^T$ for the eigenvalue λ . Then

$T\left(\mathbf{v}\right) = \rho_B^{-1}\left(\rho_B\left(T\left(\mathbf{v}\right)\right)\right)$	Definition IVLT [1214]
$=\rho_{B}^{-1}\left(M_{B,B}^{T}\rho_{B}\left(\mathbf{v}\right)\right)$	Theorem FTMR [1293]
$=\rho_{B}^{-1}\left(\lambda\rho_{B}\left(\mathbf{v}\right) \right)$	Definition EEM [945]
$=\lambda\rho_{B}^{-1}\left(\rho_{B}\left(\mathbf{v}\right)\right)$	Theorem ILTLT [1219]
$=\lambda {f v}$	Definition IVLT [1214]

which by Definition EELT [1356] says **v** is an eigenvector of T for the eigenvalue λ .

Subsection CELT Computing Eigenvectors of Linear Transformations

Knowing that the eigenvalues of a linear transformation are the eigenvalues of any representation, no matter what the choice of the basis B might be, we could now unambiguously define items such as the characteristic polynomial of a linear transformation, rather than a matrix. We'll say that again — eigenvalues, eigenvectors, and characteristic polynomials are intrinsic properties of a linear transformation, independent of the choice of a basis used to construct a matrix representation.

As a practical matter, how does one compute the eigenvalues and eigenvectors of a linear transformation of the form $T: V \to V$? Choose a nice basis B for V, one where the vector representations of the values of the linear transformations necessary for the matrix representation are easy to compute. Construct the matrix representation relative to this basis, and find the eigenvalues and eigenvectors of this matrix using the techniques of Chapter E [944]. The resulting eigenvalues of the matrix are precisely the eigenvalues of the linear transformation. The eigenvectors of the matrix are column vectors that need to be converted to vectors in V through application of ρ_B^{-1} .

Now consider the case where the matrix representation of a linear transformation is diagonalizable. The *n* linearly independent eigenvectors that must exist for the matrix (Theorem DC [1039]) can be converted (via ρ_B^{-1}) into eigenvectors of the linear transformation. A matrix representation of the linear transformation relative to a basis of eigenvectors will be a diagonal matrix — an especially nice representation! Though we did not know it at the time, the diagonalizations of Section SD [1029] were really finding especially pleasing matrix representations of linear transformations.

Here are some examples.

Example ELTT Eigenvectors of a linear transformation, twice

Consider the linear transformation $S: M_{22} \to M_{22}$ defined by

$$S\left(\begin{bmatrix}a & b\\c & d\end{bmatrix}\right) = \begin{bmatrix}-b-c-3d & -14a-15b-13c+d\\18a+21b+19c+3d & -6a-7b-7c-3d\end{bmatrix}$$

To find the eigenvalues and eigenvectors of S we will build a matrix representation and analyze the matrix. Since Theorem EER [1384] places no restriction on the choice of the basis B, we may as well use a basis that is easy to work with. So set

$$B = \{\mathbf{x}_1, \, \mathbf{x}_2, \, \mathbf{x}_3, \, \mathbf{x}_4\} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Then to build the matrix representation of S relative to B compute,

$$\rho_B \left(S \left(\mathbf{x}_1 \right) \right) = \rho_B \left(\begin{bmatrix} 0 & -14 \\ 18 & -6 \end{bmatrix} \right) = \rho_B \left(0 \mathbf{x}_1 + (-14) \mathbf{x}_2 + 18 \mathbf{x}_3 + (-6) \mathbf{x}_4 \right) = \begin{bmatrix} 0 \\ -14 \\ 18 \\ -6 \end{bmatrix}$$

$$\rho_B \left(S \left(\mathbf{x}_2 \right) \right) = \rho_B \left(\begin{bmatrix} -1 & -15\\ 21 & -7 \end{bmatrix} \right) = \rho_B \left((-1)\mathbf{x}_1 + (-15)\mathbf{x}_2 + 21\mathbf{x}_3 + (-7)\mathbf{x}_4 \right) = \begin{bmatrix} -1\\ -15\\ 21\\ -7 \end{bmatrix}$$

$$\rho_B \left(S \left(\mathbf{x}_3 \right) \right) = \rho_B \left(\begin{bmatrix} -1 & -13\\ 19 & -7 \end{bmatrix} \right) = \rho_B \left((-1)\mathbf{x}_1 + (-13)\mathbf{x}_2 + 19\mathbf{x}_3 + (-7)\mathbf{x}_4 \right) = \begin{bmatrix} -1\\ -13\\ 19\\ -7 \end{bmatrix}$$

$$\rho_B\left(S\left(\mathbf{x}_4\right)\right) = \rho_B\left(\begin{bmatrix}-3 & 1\\ 3 & -3\end{bmatrix}\right) = \rho_B\left((-3)\mathbf{x}_1 + 1\mathbf{x}_2 + 3\mathbf{x}_3 + (-3)\mathbf{x}_4\right) = \begin{bmatrix}1\\3\\-3\end{bmatrix}$$

So by Definition MR [1286] we have

$$M = M_{B,B}^S = \begin{bmatrix} 0 & -1 & -1 & -3\\ -14 & -15 & -13 & 1\\ 18 & 21 & 19 & 3\\ -6 & -7 & -7 & -3 \end{bmatrix}$$

Now compute eigenvalues and eigenvectors of the matrix representation of M with the techniques of Section EE [945]. First the characteristic polynomial,

$$p_M(x) = \det(M - xI_4) = x^4 - x^3 - 10x^2 + 4x + 24 = (x - 3)(x - 2)(x + 2)^2$$

Γ.

We could now make statements about the eigenvalues of M, but in light of Theorem EER [1384] we can refer to the eigenvalues of S and mildly abuse (or extend) our notation for multiplicities to write

$$\alpha_S(3) = 1 \qquad \qquad \alpha_S(2) = 1 \qquad \qquad \alpha_S(-2) = 2$$

Now compute the eigenvectors of M,

$$\lambda = 3 \qquad M - 3I_4 = \begin{bmatrix} -3 & -1 & -1 & -3\\ -14 & -18 & -13 & 1\\ 18 & 21 & 16 & 3\\ -6 & -7 & -7 & -6 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 1\\ 0 & 1 & 0 & -3\\ 0 & 0 & 1 & 3\\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_M(3) = \mathcal{N}(M - 3I_4) = \left\langle \left\{ \begin{bmatrix} -1\\ 3\\ -3\\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 2 \qquad M - 2I_4 = \begin{bmatrix} -2 & -1 & -1 & -3\\ -14 & -17 & -13 & 1\\ 18 & 21 & 17 & 3\\ -6 & -7 & -7 & -5 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 2\\ 0 & 1 & 0 & -4\\ 0 & 0 & 1 & 3\\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_M(2) = \mathcal{N}(M - 2I_4) = \left\langle \left\{ \begin{bmatrix} -2\\4\\-3\\1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -2 \qquad M - (-2)I_4 = \begin{bmatrix} 2 & -1 & -1 & -3\\ -14 & -13 & -13 & 1\\ 18 & 21 & 21 & 3\\ -6 & -7 & -7 & -1 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & -1\\ 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_M(-2) = \mathcal{N}(M - (-2)I_4) = \left\langle \left\{ \begin{bmatrix} 0\\ -1\\ 1\\ 0\\ \end{bmatrix}, \begin{bmatrix} 1\\ -1\\ 0\\ 1\\ \end{bmatrix} \right\} \right\rangle$$

According to Theorem EER [1384] the eigenvectors just listed as basis vectors for the eigenspaces of M are vector representations (relative to B) of eigenvectors for S. So the application if the inverse function ρ_B^{-1} will convert these column vectors into elements of the vector space M_{22} (2 × 2 matrices) that are eigenvectors of S. Since ρ_B is an isomorphism (Theorem VRILT [1272]), so is ρ_B^{-1} . Applying the inverse function will then preserve linear independence and spanning properties, so with a sweeping application of the Coordinatization Principle [1279] and some extensions of our previous notation for eigenspaces and

geometric multiplicities, we can write,

$$\rho_{B}^{-1} \left(\begin{bmatrix} -1\\3\\-3\\1 \end{bmatrix} \right) = (-1)\mathbf{x}_{1} + 3\mathbf{x}_{2} + (-3)\mathbf{x}_{3} + 1\mathbf{x}_{4} = \begin{bmatrix} -1&3\\-3&1 \end{bmatrix}$$
$$\rho_{B}^{-1} \left(\begin{bmatrix} -2\\4\\-3\\1 \end{bmatrix} \right) = (-2)\mathbf{x}_{1} + 4\mathbf{x}_{2} + (-3)\mathbf{x}_{3} + 1\mathbf{x}_{4} = \begin{bmatrix} -2&4\\-3&1 \end{bmatrix}$$
$$\rho_{B}^{-1} \left(\begin{bmatrix} 0\\-1\\1\\0 \end{bmatrix} \right) = 0\mathbf{x}_{1} + (-1)\mathbf{x}_{2} + 1\mathbf{x}_{3} + 0\mathbf{x}_{4} = \begin{bmatrix} 0&-1\\1&0 \end{bmatrix}$$
$$\rho_{B}^{-1} \left(\begin{bmatrix} 1\\-1\\0\\1 \end{bmatrix} \right) = 1\mathbf{x}_{1} + (-1)\mathbf{x}_{2} + 0\mathbf{x}_{3} + 1\mathbf{x}_{4} = \begin{bmatrix} 1&-1\\0&1 \end{bmatrix}$$

So

$$\mathcal{E}_{S}(3) = \left\langle \left\{ \begin{bmatrix} -1 & 3 \\ -3 & 1 \end{bmatrix} \right\} \right\rangle$$

$$\mathcal{E}_{S}(2) = \left\langle \left\{ \begin{bmatrix} -2 & 4 \\ -3 & 1 \end{bmatrix} \right\} \right\rangle$$
$$\mathcal{E}_{S}(-2) = \left\langle \left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right\} \right\rangle$$

with geometric multiplicities given by

$$\gamma_S(3) = 1 \qquad \qquad \gamma_S(2) = 1 \qquad \qquad \gamma_S(-2) = 2$$

Suppose we now decided to build another matrix representation of S, only now relative to a linearly independent set of eigenvectors of S, such as

$$C = \left\{ \begin{bmatrix} -1 & 3 \\ -3 & 1 \end{bmatrix}, \begin{bmatrix} -2 & 4 \\ -3 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right\}$$

At this point you should have computed enough matrix representations to predict that the result of representing S relative to C will be a diagonal matrix. Computing this representation is an example of how Theorem SCB [1377] generalizes the diagonalizations from Section SD [1029]. For the record, here is the diagonal representation,

$$M_{C,C}^{S} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

Our interest in this example is not necessarily building nice representations, but instead we want to demonstrate how eigenvalues and eigenvectors are an intrinsic property of a linear transformation, independent of any particular representation. To this end, we will repeat the foregoing, but replace B by another basis. We will make this basis different, but not extremely so,

$$D = \{\mathbf{y}_1, \, \mathbf{y}_2, \, \mathbf{y}_3, \, \mathbf{y}_4\} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\}$$

Then to build the matrix representation of S relative to D compute,

$$\rho_{D} \left(S \left(\mathbf{y}_{1} \right) \right) = \rho_{D} \left(\begin{bmatrix} 0 & -14 \\ 18 & -6 \end{bmatrix} \right) = \rho_{D} \left(14\mathbf{y}_{1} + (-32)\mathbf{y}_{2} + 24\mathbf{y}_{3} + (-6)\mathbf{y}_{4} \right) = \begin{bmatrix} 14 \\ -32 \\ 24 \\ -6 \end{bmatrix}$$

$$\rho_{D} \left(S \left(\mathbf{y}_{2} \right) \right) = \rho_{D} \left(\begin{bmatrix} -1 & -29 \\ 39 & -13 \end{bmatrix} \right) = \rho_{D} \left(28\mathbf{y}_{1} + (-68)\mathbf{y}_{2} + 52\mathbf{y}_{3} + (-13)\mathbf{y}_{4} \right) = \begin{bmatrix} 28 \\ -68 \\ 52 \\ -13 \end{bmatrix}$$

$$\rho_{D} \left(S \left(\mathbf{y}_{3} \right) \right) = \rho_{D} \left(\begin{bmatrix} -2 & -42 \\ 58 & -20 \end{bmatrix} \right) = \rho_{D} \left(40\mathbf{y}_{1} + (-100)\mathbf{y}_{2} + 78\mathbf{y}_{3} + (-20)\mathbf{y}_{4} \right) = \begin{bmatrix} 40 \\ -100 \\ 78 \\ -20 \end{bmatrix}$$

$$\rho_D\left(S\left(\mathbf{y}_4\right)\right) = \rho_D\left(\begin{bmatrix}-5 & -41\\61 & -23\end{bmatrix}\right) = \rho_D\left(36\mathbf{y}_1 + (-102)\mathbf{y}_2 + 84\mathbf{y}_3 + (-23)\mathbf{y}_4\right) = \begin{bmatrix}36\\-102\\84\\-23\end{bmatrix}$$

So by Definition MR [1286] we have

$$N = M_{D,D}^S = \begin{bmatrix} 14 & 28 & 40 & 36\\ -32 & -68 & -100 & -102\\ 24 & 52 & 78 & 84\\ -6 & -13 & -20 & -23 \end{bmatrix}$$

Now compute eigenvalues and eigenvectors of the matrix representation of N with the techniques of Section EE [945]. First the characteristic polynomial,

$$p_N(x) = \det(N - xI_4) = x^4 - x^3 - 10x^2 + 4x + 24 = (x - 3)(x - 2)(x + 2)^2$$

Of course this is not news. We now know that $M = M_{B,B}^S$ and $N = M_{D,D}^S$ are similar matrices (Theorem SCB [1377]). But Theorem SMEE [1035] told us long ago that similar matrices have identical characteristic polynomials. Now compute eigenvectors for the matrix representation, which will be different than what we found for M,

$$\lambda = 3 \qquad \qquad N - 3I_4 = \begin{bmatrix} 11 & 28 & 40 & 36 \\ -32 & -71 & -100 & -102 \\ 24 & 52 & 75 & 84 \\ -6 & -13 & -20 & -26 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -6 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_{N}(3) = \mathcal{N}(N - 3I_{4}) = \left\langle \left\{ \begin{bmatrix} -4\\6\\-4\\1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 2 \qquad N - 2I_4 = \begin{bmatrix} 12 & 28 & 40 & 36 \\ -32 & -70 & -100 & -102 \\ 24 & 52 & 76 & 84 \\ -6 & -13 & -20 & -25 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_N(2) = \mathcal{N}(N - 2I_4) = \left\langle \left\{ \begin{bmatrix} -6 \\ 7 \\ -4 \\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = -2 \qquad N - (-2)I_4 = \begin{bmatrix} 16 & 28 & 40 & 36 \\ -32 & -66 & -100 & -102 \\ 24 & 52 & 80 & 84 \\ -6 & -13 & -20 & -21 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & -1 & -3 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_N(-2) = \mathcal{N}(N - (-2)I_4) = \left\langle \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$

Employing Theorem EER [1384] we can apply ρ_D^{-1} to each of the basis vectors of the eigenspaces of N to obtain eigenvectors for S that also form bases for eigenspaces of S,

$$\rho_D^{-1} \left(\begin{bmatrix} -4\\ 6\\ -4\\ 1 \end{bmatrix} \right) = (-4)\mathbf{y}_1 + 6\mathbf{y}_2 + (-4)\mathbf{y}_3 + 1\mathbf{y}_4 = \begin{bmatrix} -1 & 3\\ -3 & 1 \end{bmatrix}$$
$$\rho_D^{-1} \left(\begin{bmatrix} -6\\ 7\\ -4\\ 1 \end{bmatrix} \right) = (-6)\mathbf{y}_1 + 7\mathbf{y}_2 + (-4)\mathbf{y}_3 + 1\mathbf{y}_4 = \begin{bmatrix} -2 & 4\\ -3 & 1 \end{bmatrix}$$
$$\rho_D^{-1} \left(\begin{bmatrix} 1\\ -2\\ 1\\ 0 \end{bmatrix} \right) = 1\mathbf{y}_1 + (-2)\mathbf{y}_2 + 1\mathbf{y}_3 + 0\mathbf{y}_4 = \begin{bmatrix} 0 & -1\\ 1 & 0 \end{bmatrix}$$
$$\rho_D^{-1} \left(\begin{bmatrix} 3\\ -3\\ 0\\ 1 \end{bmatrix} \right) = 3\mathbf{y}_1 + (-3)\mathbf{y}_2 + 0\mathbf{y}_3 + 1\mathbf{y}_4 = \begin{bmatrix} 1 & -2\\ 1 & 1 \end{bmatrix}$$

The eigenspaces for the eigenvalues of algebraic multiplicity 1 are exactly as before,

$$\mathcal{E}_{S}(3) = \left\langle \left\{ \begin{bmatrix} -1 & 3 \\ -3 & 1 \end{bmatrix} \right\} \right\rangle$$
$$\mathcal{E}_{S}(2) = \left\langle \left\{ \begin{bmatrix} -2 & 4 \\ -3 & 1 \end{bmatrix} \right\} \right\rangle$$

However, the eigenspace for $\lambda = -2$ would at first glance appear to be different. Here are the two eigenspaces for $\lambda = -2$, first the eigenspace obtained from $M = M_{B,B}^S$, then followed by the eigenspace obtained from $M = M_{D,D}^S$.

$$\mathcal{E}_{S}(-2) = \left\langle \left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right\} \right\rangle \qquad \mathcal{E}_{S}(-2) = \left\langle \left\{ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} \right\} \right\rangle$$

Subspaces generally have many bases, and that is the situation here. With a careful proof of set equality, you can show that these two eigenspaces are equal sets. The key observation to make such a proof go is that

$$\begin{bmatrix} 1 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

which will establish that the second set is a subset of the first. With equal dimensions, Theorem EDYES [858] will finish the task. So the eigenvalues of a linear transformation are independent of the matrix representation employed to compute them! \boxtimes

Another example, this time a bit larger and with complex eigenvalues.

Example CELT Complex eigenvectors of a linear transformation

Consider the linear transformation $Q: P_4 \to P_4$ defined by

$$Q(a + bx + cx^{2} + dx^{3} + ex^{4})$$

= $(-46a - 22b + 13c + 5d + e) + (117a + 57b - 32c - 15d - 4e)x + (-69a - 29b + 21c - 7e)x^{2} + (159a + 73b - 44c - 13d + 2e)x^{3} + (-195a - 87b + 55c + 10d - 13e)x^{4}$

Choose a simple basis to compute with, say

$$B = \left\{1, \, x, \, x^2, \, x^3, \, x^4\right\}$$

Then it should be apparent that the matrix representation of Q relative to B is

$$M = M_{B,B}^Q = \begin{bmatrix} -46 & -22 & 13 & 5 & 1\\ 117 & 57 & -32 & -15 & -4\\ -69 & -29 & 21 & 0 & -7\\ 159 & 73 & -44 & -13 & 2\\ -195 & -87 & 55 & 10 & -13 \end{bmatrix}$$

Compute the characteristic polynomial, eigenvalues and eigenvectors according to the techniques of Section EE [945],

$$p_Q(x) = -x^5 + 6x^4 - x^3 - 88x^2 + 252x - 208$$

$$= -(x-2)^{2}(x+4) (x^{2}-6x+13)$$

= -(x-2)^{2}(x+4) (x - (3+2i)) (x - (3-2i))

$$\alpha_Q(2) = 2$$
 $\alpha_Q(-4) = 1$ $\alpha_Q(3+2i) = 1$ $\alpha_Q(3-2i) = 1$

$$\lambda = 2$$

$$M - (2)I_5 = \begin{bmatrix} -48 & -22 & 13 & 5 & 1\\ 117 & 55 & -32 & -15 & -4\\ -69 & -29 & 19 & 0 & -7\\ 159 & 73 & -44 & -15 & 2\\ -195 & -87 & 55 & 10 & -15 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2}\\ 0 & 1 & 0 & -\frac{5}{2} & -\frac{5}{2}\\ 0 & 0 & 1 & -2 & -6\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_M(2) = \mathcal{N}(M - (2)I_5) = \left\langle \left\{ \begin{bmatrix} -\frac{1}{2}\\ \frac{5}{2}\\ 2\\ 1\\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{2}\\ \frac{5}{2}\\ 2\\ 1\\ 0 \end{bmatrix}, \left\{ \begin{bmatrix} \frac{1}{2}\\ \frac{5}{2}\\ 2\\ 1\\ 0 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -1\\ 5\\ 4\\ 2\\ 0 \end{bmatrix}, \begin{bmatrix} 1\\ 5\\ 12\\ 0\\ 2 \end{bmatrix} \right\} \right\rangle$$

 $\lambda = -4$

$$M - (-4)I_{5} = \begin{bmatrix} -42 & -22 & 13 & 5 & 1\\ 117 & 61 & -32 & -15 & -4\\ -69 & -29 & 25 & 0 & -7\\ 159 & 73 & -44 & -9 & 2\\ -195 & -87 & 55 & 10 & -9 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & -3\\ 0 & 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 1 & -2\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathcal{E}_{M}(-4) = \mathcal{N}(M - (-4)I_{5}) = \left\langle \left\{ \begin{bmatrix} -1\\ 3\\ 1\\ 2\\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 3 + 2i$$

$$M - (3 + 2i)I_5 = \begin{bmatrix} -49 - 2i & -22 & 13 & 5 & 1\\ 117 & 54 - 2i & -32 & -15 & -4\\ -69 & -29 & 18 - 2i & 0 & -7\\ 159 & 73 & -44 & -16 - 2i & 2\\ -195 & -87 & 55 & 10 & -16 - 2i \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{3}{4} + \frac{i}{4} \\ 0 & 1 & 0 & 0 & \frac{7}{4} - \frac{i}{4} \\ 0 & 0 & 1 & 0 & -\frac{1}{2} + \frac{i}{2} \\ 0 & 0 & 0 & 1 & \frac{7}{4} - \frac{i}{4} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_M (3 + 2i) = \mathcal{N}(M - (3 + 2i)I_5) = \left\langle \left\{ \begin{bmatrix} \frac{3}{4} - \frac{i}{4} \\ -\frac{7}{4} + \frac{i}{4} \\ \frac{1}{2} - \frac{i}{2} \\ -\frac{7}{4} + \frac{i}{4} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 3 - i \\ -7 + i \\ 2 - 2i \\ -7 + i \\ 4 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 3 - 2i$$

$$M - (3 - 2i)I_5 = \begin{bmatrix} -49 + 2i & -22 & 13 & 5 & 1\\ 117 & 54 + 2i & -32 & -15 & -4\\ -69 & -29 & 18 + 2i & 0 & -7\\ 159 & 73 & -44 & -16 + 2i & 2\\ -195 & -87 & 55 & 10 & -16 + 2i \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{3}{4} - \frac{i}{4} \\ 0 & 1 & 0 & 0 & \frac{7}{4} + \frac{i}{4} \\ 0 & 0 & 1 & 0 & -\frac{1}{2} - \frac{i}{2} \\ 0 & 0 & 0 & 1 & \frac{7}{4} + \frac{i}{4} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_M (3 - 2i) = \mathcal{N}(M - (3 - 2i)I_5) = \left\langle \left\{ \begin{bmatrix} \frac{3}{4} + \frac{i}{4} \\ -\frac{7}{4} - \frac{i}{4} \\ \frac{1}{2} + \frac{i}{2} \\ -\frac{7}{4} - \frac{i}{4} \\ 1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \begin{bmatrix} 3 + i \\ -7 - i \\ 2 + 2i \\ -7 - i \\ 4 \end{bmatrix} \right\} \right\rangle$$

It is straightforward to convert each of these basis vectors for eigenspaces of M back to elements of P_4 by applying the isomorphism ρ_B^{-1} ,

$$\rho_B^{-1} \left(\begin{bmatrix} -1\\5\\4\\2\\0 \end{bmatrix} \right) = -1 + 5x + 4x^2 + 2x^3$$

$$\begin{split} \rho_B^{-1} \left(\begin{bmatrix} 1\\5\\12\\0\\2 \end{bmatrix} \right) &= 1 + 5x + 12x^2 + 2x^4 \\ \rho_B^{-1} \left(\begin{bmatrix} -1\\3\\1\\2\\1 \end{bmatrix} \right) &= -1 + 3x + x^2 + 2x^3 + x^4 \\ \rho_B^{-1} \left(\begin{bmatrix} 3-i\\-7+i\\2-2i\\-7+i\\4 \end{bmatrix} \right) &= (3-i) + (-7+i)x + (2-2i)x^2 + (-7+i)x^3 + 4x^4 \\ \rho_B^{-1} \left(\begin{bmatrix} 3+i\\-7-i\\2+2i\\-7-i\\4 \end{bmatrix} \right) &= (3+i) + (-7-i)x + (2+2i)x^2 + (-7-i)x^3 + 4x^4 \end{split}$$

So we apply Theorem EER [1384] and the Coordinatization Principle [1279] to get the eigenspaces for Q,

$$\mathcal{E}_Q(2) = \left\langle \left\{ -1 + 5x + 4x^2 + 2x^3, 1 + 5x + 12x^2 + 2x^4 \right\} \right\rangle$$

$$\mathcal{E}_Q(-4) = \left\langle \left\{ -1 + 3x + x^2 + 2x^3 + x^4 \right\} \right\rangle$$

$$\mathcal{E}_Q(3 + 2i) = \left\langle \left\{ (3 - i) + (-7 + i)x + (2 - 2i)x^2 + (-7 + i)x^3 + 4x^4 \right\} \right\rangle$$

$$\mathcal{E}_Q(3 - 2i) = \left\langle \left\{ (3 + i) + (-7 - i)x + (2 + 2i)x^2 + (-7 - i)x^3 + 4x^4 \right\} \right\rangle$$

with geometric multiplicities

$$\gamma_Q(2) = 2$$
 $\gamma_Q(-4) = 1$ $\gamma_Q(3+2i) = 1$ $\gamma_Q(3-2i) = 1$

Subsection READ Reading Questions

- 1. The change-of-basis matrix is a matrix representation of which linear transformation?
- 2. Find the change-of-basis matrix, $C_{B,C}$, for the two bases of \mathbb{C}^2

$$B = \left\{ \begin{bmatrix} 2\\3 \end{bmatrix}, \begin{bmatrix} -1\\2 \end{bmatrix} \right\} \qquad \qquad C = \left\{ \begin{bmatrix} 1\\0 \end{bmatrix}, \begin{bmatrix} 1\\1 \end{bmatrix} \right\}$$

 \boxtimes

3. What is the third "surprise," and why is it surprising?

Subsection EXC Exercises

C20 In Example CBCV [1367] we computed the vector representation of \mathbf{y} relative to C, $\rho_C(\mathbf{y})$, as an example of Theorem CB [1359]. Compute this same representation directly. In other words, apply Definition VR [1261] rather than Theorem CB [1359]. Contributed by Robert Beezer

C21 Perform a check on Example MRCM [1372] by computing $M_{B,D}^Q$ directly. In other words, apply Definition MR [1286] rather than Theorem MRCB [1371]. Contributed by Robert Beezer Solution [1406]

C30 Find a basis for the vector space P_3 composed of eigenvectors of the linear transformation T. Then find a matrix representation of T relative to this basis.

$$T: P_3 \to P_3, \quad T\left(a + bx + cx^2 + dx^3\right) = (a + c + d) + (b + c + d)x + (a + b + c)x^2 + (a + b + d)x^3$$

Contributed by Robert Beezer Solution [1407]

C40 Let S_{22} be the vector space of 2×2 symmetric matrices. Find a basis B for S_{22} that yields a diagonal matrix representation of the linear transformation R.

$$R: S_{22} \to S_{22}, \quad R\left(\begin{bmatrix}a & b\\ b & c\end{bmatrix}\right) = \begin{bmatrix}-5a+2b-3c & -12a+5b-6c\\ -12a+5b-6c & 6a-2b+4c\end{bmatrix}$$

Contributed by Robert Beezer Solution [1409]

C41 Let S_{22} be the vector space of 2×2 symmetric matrices. Find a basis for S_{22} composed of eigenvectors of the linear transformation $Q: S_{22} \to S_{22}$.

$$Q\left(\begin{bmatrix} a & b \\ b & c \end{bmatrix}\right) = \begin{bmatrix} 25a + 18b + 30c & -16a - 11b - 20c \\ -16a - 11b - 20c & -11a - 9b - 12c \end{bmatrix}$$

Contributed by Robert Beezer Solution [1411]

T10 Suppose that $T: V \to V$ is an invertible linear transformation with a nonzero eigenvalue λ . Prove that $\frac{1}{\lambda}$ is an eigenvalue of T^{-1} . Contributed by Robert Beezer Solution [1412]

T15 Suppose that V is a vector space and $T: V \to V$ is a linear transformation. Prove that T is injective if and only if $\lambda = 0$ is not an eigenvalue of T. Contributed by Robert Beezer

Subsection SOL Solutions

C21 Contributed by Robert Beezer Statement [1404] Apply Definition MR [1286],

$$\begin{split} \rho_D \left(Q \left(\begin{bmatrix} 5 & -3 \\ -3 & -2 \end{bmatrix} \right) \right) &= \rho_D \left(19 + 14x - 2x^2 - 28x^3 \right) \\ &= \rho_D \left((-39)(2 + x - 2x^2 + 3x^3) + 62(-1 - 2x^2 + 3x^3) + (-53)(-3 - x + x^3) + (-44)(-x^2 + x^3) \right) \\ &= \begin{bmatrix} -39 \\ 62 \\ -53 \\ -44 \end{bmatrix} \\ \rho_D \left(Q \left(\begin{bmatrix} 2 & -3 \\ -3 & 0 \end{bmatrix} \right) \right) &= \rho_D \left(16 + 9x - 7x^2 - 14x^3 \right) \\ &= \rho_D \left((-23)(2 + x - 2x^2 + 3x^3) + (34)(-1 - 2x^2 + 3x^3) + (-32)(-3 - x + x^3) + (-15)(-x^2 + x^3) \right) \\ &= \begin{bmatrix} -23 \\ 34 \\ -32 \\ -15 \end{bmatrix} \end{split}$$

$$\rho_D \left(Q \left(\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \right) \right) = \rho_D \left(25 + 9x + 3x^2 + 4x^3 \right)$$

= $\rho_D \left((14)(2 + x - 2x^2 + 3x^3) + (-12)(-1 - 2x^2 + 3x^3) + 5(-3 - x + x^3) + (-7)(-x^2 + x^3) \right)$
= $\begin{bmatrix} 14 \\ -12 \\ 5 \\ -7 \end{bmatrix}$

These three vectors are the columns of the matrix representation,

$$M_{B,D}^Q = \begin{bmatrix} -39 & -23 & 14\\ 62 & 34 & -12\\ -53 & -32 & 5\\ -44 & -15 & -7 \end{bmatrix}$$

which coincides with the result obtained in Example MRCM [1372].

C30 Contributed by Robert Beezer Statement [1404]

With the domain and codomain being identical, we will build a matrix representation using the same basis for both the domain and codomain. The eigenvalues of the matrix representation will be the eigenvalues of the linear transformation, and we can obtain the eigenvectors of the linear transformation by un-coordinatizing (Theorem EER [1384]). Since the method does not depend on *which* basis we choose, we can choose a natural basis for ease of computation, say,

$$B = \left\{1, \, x, \, x^2, x^3\right\}$$

The matrix representation is then,

$$M_{B,B}^{T} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

The eigenvalues and eigenvectors of this matrix were computed in Example ESMS4 [970]. A basis for \mathbb{C}^4 , composed of eigenvectors of the matrix representation is,

$$C = \left\{ \begin{bmatrix} 1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} -1\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\-1\\1 \end{bmatrix}, \begin{bmatrix} -1\\-1\\1\\1 \end{bmatrix} \right\}$$

Applying ρ_B^{-1} to each vector of this set, yields a basis of P_3 composed of eigenvectors of T,

$$D = \left\{1 + x + x^2 + x^3, -1 + x, \ -x^2 + x^3, \ -1 - x + x^2 + x^3\right\}$$

The matrix representation of T relative to the basis D will be a diagonal matrix with the

corresponding eigenvalues along the diagonal, so in this case we get

$$M_{D,D}^{T} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

C40 Contributed by Robert Beezer Statement [1404]

Begin with a matrix representation of R, any matrix representation, but use the same basis for both instances of S_{22} . We'll choose a basis that makes it easy to compute vector representations in S_{22} .

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Then the resulting matrix representation of R (Definition MR [1286]) is

$$M_{B,B}^{R} = \begin{bmatrix} -5 & 2 & -3\\ -12 & 5 & -6\\ 6 & -2 & 4 \end{bmatrix}$$

Now, compute the eigenvalues and eigenvectors of this matrix, with the goal of diagonalizing the matrix (Theorem DC [1039]),

$$\lambda = 2 \qquad \qquad \mathcal{E}_{M_{B,B}^{R}}\left(2\right) = \left\langle \left\{ \begin{bmatrix} -1\\ -2\\ 1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 1 \qquad \qquad \mathcal{E}_{M_{B,B}^{R}}\left(1\right) = \left\langle \left\{ \begin{bmatrix} -1\\0\\2 \end{bmatrix}, \begin{bmatrix} 1\\3\\0 \end{bmatrix} \right\} \right\rangle$$

The three vectors that occur as basis elements for these eigenspaces will together form a linearly independent set (check this!). So these column vectors may be employed in a matrix that will diagonalize the matrix representation. If we "un-coordinatize" these three column vectors relative to the basis B, we will find three linearly independent elements of S_{22} that are eigenvectors of the linear transformation R (Theorem EER [1384]). A matrix representation relative to this basis of eigenvectors will be diagonal, with the eigenvalues ($\lambda = 2, 1$) as the diagonal elements. Here we go,

$$\rho_B^{-1} \left(\begin{bmatrix} -1\\ -2\\ 1 \end{bmatrix} \right) = (-1) \begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + (-2) \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 1 \begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & -2\\ -2 & 1 \end{bmatrix}$$
$$\rho_B^{-1} \left(\begin{bmatrix} -1\\ 0\\ 2 \end{bmatrix} \right) = (-1) \begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + 0 \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 2 \begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0\\ 0 & 2 \end{bmatrix}$$
$$\rho_B^{-1} \left(\begin{bmatrix} 1\\ 3\\ 0 \end{bmatrix} \right) = 1 \begin{bmatrix} 1 & 0\\ 0 & 0 \end{bmatrix} + 3 \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix} + 0 \begin{bmatrix} 0 & 0\\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3\\ 3 & 0 \end{bmatrix}$$

So the requested basis of S_{22} , yielding a diagonal matrix representation of R, is

$$\left\{ \begin{bmatrix} -1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 3 & 0 \end{bmatrix} \right\}$$

C41 Contributed by Robert Beezer Statement [1405] Use a single basis for both the domain and codomain, since they are equal.

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

The matrix representation of Q relative to B is

$$M = M_{B,B}^Q = \begin{bmatrix} 25 & 18 & 30\\ -16 & -11 & -20\\ -11 & -9 & -12 \end{bmatrix}$$

We can analyze this matrix with the techniques of Section EE [945] and then apply Theorem EER [1384]. The eigenvalues of this matrix are $\lambda = -2$, 1, 3 with eigenspaces

$$\mathcal{E}_{M}(-2) = \left\langle \left\{ \begin{bmatrix} -6\\4\\3 \end{bmatrix} \right\} \right\rangle \qquad \mathcal{E}_{M}(1) = \left\langle \left\{ \begin{bmatrix} -2\\1\\1 \end{bmatrix} \right\} \right\rangle \qquad \mathcal{E}_{M}(3) = \left\langle \left\{ \begin{bmatrix} -3\\2\\1 \end{bmatrix} \right\} \right\rangle$$

Because the three eigenvalues are distinct, the three basis vectors from the three eigenspaces for a linearly independent set (Theorem EDELI [1001]). Theorem EER [1384] says we can

uncoordinatize these eigenvectors to obtain eigenvectors of Q. By Theorem ILTLI [1148] the resulting set will remain linearly independent. Set

$$C = \left\{ \rho_B^{-1} \left(\begin{bmatrix} -6\\4\\3 \end{bmatrix} \right), \rho_B^{-1} \left(\begin{bmatrix} -2\\1\\1 \end{bmatrix} \right), \rho_B^{-1} \left(\begin{bmatrix} -3\\2\\1 \end{bmatrix} \right) \right\} = \left\{ \begin{bmatrix} -6&4\\4&3 \end{bmatrix}, \begin{bmatrix} -2&1\\1&1 \end{bmatrix}, \begin{bmatrix} -3&2\\2&1 \end{bmatrix} \right\}$$

Then C is a linearly independent set of size 3 in the vector space S_{22} , which has dimension 3 as well. By Theorem G [852], C is a basis of S_{22} .

T10 Contributed by Robert Beezer Statement [1405] Let \mathbf{v} be an eigenvector of T for the eigenvalue λ . Then,

$$T^{-1} (\mathbf{v}) = \frac{1}{\lambda} \lambda T^{-1} (\mathbf{v}) \qquad \lambda \neq 0$$

= $\frac{1}{\lambda} T^{-1} (\lambda \mathbf{v}) \qquad \text{Theorem ILTLT [1219]}$
= $\frac{1}{\lambda} T^{-1} (T (\mathbf{v})) \qquad \mathbf{v} \text{ eigenvector of } T$
= $\frac{1}{\lambda} I_V (\mathbf{v}) \qquad \text{Definition IVLT [1214]}$
= $\frac{1}{\lambda} \mathbf{v} \qquad \text{Definition IDLT [1213]}$

which says that $\frac{1}{\lambda}$ is an eigenvalue of T^{-1} with eigenvector **v**. Note that it is possible to prove

that any eigenvalue of an invertible linear transformation is never zero. So the hypothesis that λ be nonzero is just a convenience for this problem.

Section OD Orthonormal Diagonalization

This section is in draft form Theorems & definitions are complete, needs examples Inner product is "reversed" from prior material, see changelog explanation

We have seen in Section SD [1029] that under the right conditions a square matrix is similar to a diagonal matrix. We recognize now, via Theorem SCB [1377], that a similarity transformation is a change of basis on a matrix representation. So we can now discuss the choice of a basis used to build a matrix representation, and decide if some bases are better than others for this purpose. This will be the tone of this section. We will also see that every matrix has a reasonably useful matrix representation, and we will discover a new class of diagonalizable linear transformations. First we need some basic facts about triangular matrices.

Subsection TM Triangular Matrices

An upper, or lower, triangular matrix is exactly what it sounds like it should be, but here are the two relevant definitions.

Definition UTM Upper Triangular Matrix The $n \times n$ square matrix A is **upper triangular** if $[A]_{ij} = 0$ whenever i > j.

Definition LTM Lower Triangular Matrix The $n \times n$ square matrix A is **lower triangular** if $[A]_{ij} = 0$ whenever i < j.

Obviously, properties of a lower triangular matrices will have analogues for upper triangular matrices. Rather than stating two very similar theorems, we will say that matrices are "triangular of the same type" as a convenient shorthand to cover both possibilities and then give a proof for just one type.

Theorem PTMT Product of Triangular Matrices is Triangular

Suppose that A and B are square matrices of size n that are triangular of the same type.

Then AB is also triangular of that type.

Proof We prove this for lower triangular matrices and leave the proof for upper triangular matrices to you. Suppose that A and B are both lower triangular. We need only establish that certain entries of the product AB are zero. Suppose that i < j, then

$$\begin{aligned} AB]_{ij} &= \sum_{k=1}^{n} [A]_{ik} [B]_{kj} & \text{Theorem EMP [471]} \\ &= \sum_{k=1}^{j-1} [A]_{ik} [B]_{kj} + \sum_{k=j}^{n} [A]_{ik} [B]_{kj} & \text{Property AACN [1597]} \\ &= \sum_{k=1}^{j-1} [A]_{ik} 0 + \sum_{k=j}^{n} [A]_{ik} [B]_{kj} & k < j, \text{ Definition LTM [1415]} \\ &= \sum_{k=1}^{j-1} [A]_{ik} 0 + \sum_{k=j}^{n} 0 [B]_{kj} & i < j \le k, \text{ Definition LTM [1415]} \\ &= \sum_{k=1}^{j-1} 0 + \sum_{k=j}^{n} 0 \\ &= 0 \end{aligned}$$

Since $[AB]_{ij} = 0$ whenever i < j, by Definition LTM [1415], AB is lower triangular. The inverse of a triangular matrix is triangular, of the same type. П

Theorem ITMT Inverse of a Triangular Matrix is Triangular

Suppose that A is a nonsingular matrix of size n that is triangular. Then the inverse of A, A^{-1} , is triangular of the same type. Furthermore, the diagonal entries of A^{-1} are the reciprocals of the corresponding diagonal entries of A. More precisely, $[A^{-1}]_{ii} = [A]_{ii}^{-1}$. \Box

Proof We give the proof for the case when A is lower triangular, and leave the case when A is upper triangular for you. Consider the process for computing the inverse of a matrix that is outlined in the proof of Theorem CINM [516]. We augment A with the size n identity matrix, I_n , and row-reduce the $n \times 2n$ matrix to reduced row-echelon form via the algorithm in Theorem REMEF [66]. The proof involves tracking the peculiarities of this process in the case of a lower triangular matrix. Let $M = [A | I_n]$.

First, none of the diagonal elements of A are zero. By repeated expansion about the first row, the determinant of a lower triangular matrix can be seen to be the product of the diagonal entries (Theorem DER [897]). If just one of these diagonal elements was zero, then the determinant of A is zero and A is singular by Theorem SMZD [935]. Slightly violating the exact algorithm for row reduction we can form a matrix, M', that is row-equivalent to M, by multiplying row i by the nonzero scalar $[A]_{ii}^{-1}$, for $1 \le i \le n$. This sets $[M']_{ii} = 1$ and $[M']_{i,n+1} = [A]_{ii}^{-1}$, and leaves every zero entry of M unchanged.

Let M_j denote the matrix obtained form M' after converting column j to a pivot column. We can convert column j of M_{j-1} into a pivot column with a set of n-j-1 row operations of the form $\alpha R_j + R_k$ with $j+1 \leq k \leq n$. The key observation here is that we add multiples of row j only to higher-numbered rows. This means that none of the entries in rows 1 through j-1 is changed, and since row j has zeros in columns j+1 through n, none of the entries in rows j+1 through n is changed in columns j+1 through n. The first n columns of M'form a lower triangular matrix with 1's on the diagonal. In its conversion to the identity matrix through this sequence of row operations, it remains lower triangular with 1's on the diagonal.

What happens in columns n + 1 through 2n of M'? These columns began in M as the identity matrix, and in M' each diagonal entry was scaled to a reciprocal of the corresponding diagonal entry of A. Notice that trivially, these final n columns of M' form a lower triangular matrix. Just as we argued for the first n columns, the row operations that convert M_{j-1} into M_j will preserve the lower triangular form in the final n columns and preserve the exact values of the diagonal entries. By Theorem CINM [516], the final n columns of M_n is the inverse of A, and this matrix has the necessary properties advertised in the conclusion of this theorem.

Subsection UTMR Upper Triangular Matrix Representation

Not every matrix is diagonalizable, but every linear transformation has a matrix representation that is an upper triangular matrix, and the basis that achieves this representation is especially pleasing. Here's the theorem.

Theorem UTMR

Upper Triangular Matrix Representation

Suppose that $T: V \to V$ is a linear transformation. Then there is a basis B for V such that the matrix representation of T relative to B, $M_{B,B}^T$, is an upper triangular matrix. Each diagonal entry is an eigenvalue of T, and if λ is an eigenvalue of T, then λ occurs $\alpha_T(\lambda)$ times on the diagonal.

Proof We begin with a proof by induction (Technique I [1630]) of the first statement in the conclusion of the theorem. We use induction on the dimension of V to show that if $T: V \to V$ is a linear transformation, then there is a basis B for V such that the matrix representation of T relative to $B, M_{B,B}^T$, is an upper triangular matrix.

To start suppose that dim (V) = 1. Choose any nonzero vector $\mathbf{v} \in V$ and realize that $V = \langle \{\mathbf{v}\} \rangle$. Then we can determine T uniquely by $T(\mathbf{v}) = \beta \mathbf{v}$ for some $\beta \in \mathbb{C}$ (Theorem LTDB [1094]). This description of T also gives us a matrix representation relative to the basis $B = \{\mathbf{v}\}$ as the 1×1 matrix with lone entry equal to β . And this matrix representation is upper triangular (Definition UTM [1415]).

For the induction step let dim (V) = m, and assume the theorem is true for every linear transformation defined on a vector space of dimension less than m. By Theorem EMHE [953] (suitably converted to the setting of a linear transformation), T has at least one eigenvalue, and we denote this eigenvalue as λ . (We will remark later about how critical this step is.) We now consider properties of the linear transformation $T - \lambda I_V : V \to V$.

Let **x** be an eigenvector of T for λ . By definition $\mathbf{x} \neq \mathbf{0}$. Then

 $(T - \lambda I_V)(\mathbf{x}) = T(\mathbf{x}) - \lambda I_V(\mathbf{x})$ Theorem VSLT [1110]

$=T\left(\mathbf{x}\right) -\lambda\mathbf{x}$	Definition IDLT [1213]
$=\lambda \mathbf{x} - \lambda \mathbf{x}$	Definition EELT [1356]
= 0	Property AI [660]

So $T - \lambda I_V$ is not injective, as it has a nontrivial kernel (Theorem KILT [1144]). With an application of Theorem RPNDD [1235] we bound the rank of $T - \lambda I_V$,

$$r(T - \lambda I_V) = \dim(V) - n(T - \lambda I_V) \le m - 1$$

Define W to be the subspace of V that is the range of $T - \lambda I_V$, $W = \mathcal{R}(T - \lambda I_V)$. We define a new linear transformation S, on W,

$$S \colon W \to W \qquad \qquad S\left(\mathbf{w}\right) = T\left(\mathbf{w}\right)$$

This does not look we have accomplished much, since the action of S is identical to the action of T. For our purposes this will be a good thing. What is different is the domain and codomain. S is defined on W, a vector space with dimension less than m, and so is susceptible to our induction hypothesis. Verifying that S is really a linear transformation is almost entirely routine, with one exception. Employing T in our definition of S raises the possibility that the outputs of S will not be contained within W (but instead will lie inside V, but outside W). To examine this possibility, suppose that $\mathbf{w} \in W$.

$$S(\mathbf{w}) = T(\mathbf{w})$$

= T(\mathbf{w}) + \mathbf{0} Property Z [660]

$=T\left(\mathbf{w}\right)+\left(\lambda I_{V}\left(\mathbf{w}\right)-\lambda I_{V}\left(\mathbf{w}\right)\right)$	Property AI [660]
$=\left(T\left(\mathbf{w}\right)-\lambda I_{V}\left(\mathbf{w}\right)\right)+\lambda I_{V}\left(\mathbf{w}\right)$	Property AA [660]
$=\left(T\left(\mathbf{w}\right)-\lambda I_{V}\left(\mathbf{w}\right)\right)+\lambda\mathbf{w}$	Definition IDLT [1213]
$=\left(T-\lambda I_{V}\right)\left(\mathbf{w}\right)+\lambda\mathbf{w}$	Theorem VSLT [1110]

Since W is the range of $T - \lambda I_V$, $(T - \lambda I_V)$ (\mathbf{w}) $\in W$. And by Property SC [660], $\lambda \mathbf{w} \in W$. Finally, applying Property AC [660] we see by closure that the sum is in W and so we conclude that $S(\mathbf{w}) \in W$. This argument convinces us that it is legitimate to define S as we did with W as the codomain.

S is a linear transformation defined on a vector space with dimension less than m, so we can apply the induction hypothesis and conclude that W has a basis, $C = {\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_k}$, such that the matrix representation of S relative to C is an upper triangular matrix.

By Theorem DSFOS [867] there exists a second subspace of V, which we will call U, so that V is a direct sum of W and U, $V = W \oplus U$. Choose a basis $D = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_\ell\}$ for U. So $m = k + \ell$ by Theorem DSD [872], and $B = C \cup D$ is basis for V by Theorem DSLI [870] and Theorem G [852]. B is the basis we desire. What does a matrix representation of T look like, relative to B?

Since the definition of T and S agree on W, the first k columns of $M_{B,B}^T$ will have the upper triangular matrix representation of S in the first k rows. The remaining $\ell = m - k$ rows of these first k columns will be all zeros since the outputs of T on C are all contained in W. The situation for T on D is not quite as pretty, but it is close.

For $1 \leq i \leq \ell$, consider $\rho_B\left(T\left(\mathbf{u}_i\right)\right) = \rho_B\left(T\left(\mathbf{u}_i\right) + \mathbf{0}\right)$ Property Z [660] $= \rho_B \left(T \left(\mathbf{u}_i \right) + \left(\lambda I_V \left(\mathbf{u}_i \right) - \lambda I_V \left(\mathbf{u}_i \right) \right) \right)$ Property AI [660] $= \rho_B \left(\left(T \left(\mathbf{u}_i \right) - \lambda I_V \left(\mathbf{u}_i \right) \right) + \lambda I_V \left(\mathbf{u}_i \right) \right)$ Property AA [660] $= \rho_B \left(\left(T \left(\mathbf{u}_i \right) - \lambda I_V \left(\mathbf{u}_i \right) \right) + \lambda \mathbf{u}_i \right)$ Definition IDLT [1213] $= \rho_B \left(\left(T - \lambda I_V \right) \left(\mathbf{u}_i \right) + \lambda \mathbf{u}_i \right)$ Theorem VSLT [1110] $= \rho_B \left(a_1 \mathbf{w}_1 + a_2 \mathbf{w}_2 + a_3 \mathbf{w}_3 + \dots + a_k \mathbf{w}_k + \lambda \mathbf{u}_i \right)$ Definition RLT [1178] a_2 \vdots a_k 0Definition VR [1261] 0

In the penultimate step of this proof, we have rewritten an element of the range of $T - \lambda I_V$ as a linear combination of the basis vectors, C, for the range of $T - \lambda I_V$, W, using the scalars $a_1, a_2, a_3, \ldots, a_k$. If we incorporate these ℓ column vectors into the matrix representation $M_{B,B}^T$ we find ℓ occurrences of λ on the diagonal, and any nonzero entries lying only in the first k rows. Together with the $k \times k$ upper triangular representation in the upper left-hand corner, the entire matrix representation is now clearly upper triangular. This completes the induction step, so for any linear transformation there is a basis that creates an upper triangular matrix representation.

We have one more statement in the conclusion of the theorem to verify. The eigenvalues of T, and their multiplicities, can be computed with the techniques of Chapter E [944] relative to any matrix representation (Theorem EER [1384]). We take this approach with our upper triangular matrix representation $M_{B,B}^T$. Let d_i be the diagonal entry of $M_{B,B}^T$ in row i and column i. Then the characteristic polynomial, computed as a determinant (Definition CP [960]) with repeated expansions about the first column, is

$$p_{M_{B,B}^{T}}(x) = (d_{1} - x) (d_{2} - x) (d_{3} - x) \cdots (d_{m} - x)$$

The roots of the polynomial equation $p_{M_{B,B}^T}(x) = 0$ are the eigenvalues of the linear transformation (Theorem EMRCP [962]). So each diagonal entry is an eigenvalue, and is repeated on the diagonal exactly $\alpha_T(\lambda)$ times (Definition AME [968]).

A key step in this proof was the construction of the subspace W with dimension strictly less than that of V. This required an eigenvalue/eigenvector pair, which was guaranteed to us by Theorem EMHE [953]. Digging deeper, the proof of Theorem EMHE [953] requires that we can factor polynomials completely, into linear factors. This will not always happen if our set of scalars is the reals, \mathbb{R} . So this is our final explanation of our choice of the complex numbers, \mathbb{C} , as our set of scalars. In \mathbb{C} polynomials factor completely, so every matrix has at least one eigenvalue, and an inductive argument will get us to upper triangular matrix representations.

In the case of linear transformations defined on \mathbb{C}^m , we can use the inner product (Definition IP [398]) profitably to fine-tune the basis that yields an upper triangular matrix representation. Recall that the adjoint of matrix A (Definition A [442]) is written as A^* .

Theorem OBUTR Orthonormal Basis for Upper Triangular I

Orthonormal Basis for Upper Triangular Representation

Suppose that A is a square matrix. Then there is a unitary matrix U, and an upper triangular matrix T, such that

$$U^*AU = T$$

and T has the eigenvalues of A as the entries of the diagonal.

Proof This theorem is a statement about matrices and similarity. We can convert it to a statement about linear transformations, matrix representations and bases (Theorem SCB [1377]). Suppose that A is an $n \times n$ matrix, and define the linear transformation $S: \mathbb{C}^n \to \mathbb{C}^n$ by $S(\mathbf{x}) = A\mathbf{x}$. Then Theorem UTMR [1419] gives us a basis $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_n}$ for \mathbb{C}^n such that a matrix representation of S relative to B, $M_{B,B}^S$, is upper triangular.

Now convert the basis B into an orthogonal basis, C, by an application of the Gram-Schmidt procedure (Theorem GSP [413]). This is a messy business computationally, but here we have an excellent illustration of the power of the Gram-Schmidt procedure. We need only be sure that B is linearly independent and spans \mathbb{C}^n , and then we know that

 \square

C is linearly independent, spans \mathbb{C}^n and is also an orthogonal set. We will now consider the matrix representation of *S* relative to *C* (rather than *B*). Write the new basis as $C = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_n\}$. The application of the Gram-Schmidt procedure creates each vector of *C*, say \mathbf{y}_j , as the difference of \mathbf{v}_j and a linear combination of $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_{j-1}$. We are not concerned here with the actual values of the scalars in this linear combination, so we will write

$$\mathbf{y}_j = \mathbf{v}_j - \sum_{k=1}^{j-1} b_{jk} \mathbf{y}_k$$

where the b_{jk} are shorthand for the scalars. The equation above is in a form useful for creating the basis C from B. To better understand the relationship between B and C convert it to read

$$\mathbf{v}_j = \mathbf{y}_j + \sum_{k=1}^{j-1} b_{jk} \mathbf{y}_k$$

In this form, we recognize that the change-of-basis matrix $C_{B,C} = M_{B,C}^{I_{\mathbb{C}^n}}$ (Definition CBM [1359]) is an upper triangular matrix. By Theorem SCB [1377] we have

$$M_{C,C}^S = C_{B,C} M_{B,B}^S C_{B,C}^{-1}$$

The inverse of an upper triangular matrix is upper triangular (Theorem ITMT [1417]), and the product of two upper triangular matrices is again upper triangular (Theorem PTMT [1415]). So $M_{C,C}^S$ is an upper triangular matrix. Now, multiply each vector of C by a nonzero scalar, so that the result has norm 1. In this way we create a new basis D which is an orthonormal set (Definition ONS [418]). Note that the change-of-basis matrix $C_{C,D}$ is a diagonal matrix with nonzero entries equal to the norms of the vectors in C.

Now we can convert our results into the language of matrices. Let E be the basis of \mathbb{C}^n formed with the standard unit vectors (Definition SUV [409]). Then the matrix representation of S relative to E is simply A, $A = M_{E,E}^S$. The change-of-basis matrix $C_{D,E}$ has columns that are simply the vectors in D, the orthonormal basis. As such, Theorem CUMOS [547] tells us that $C_{D,E}$ is a unitary matrix, and by Definition UM [544] has an inverse equal to its adjoint. Write $U = C_{D,E}$. We have

$$U^*AU = U^{-1}AU$$

$$= C_{D,E}^{-1}M_{E,E}^S C_{D,E}$$

$$= M_{D,D}^S$$
Theorem SCB [1377]
$$= C_{C,D}M_{C,C}^S C_{C,D}^{-1}$$
Theorem SCB [1377]

The inverse of a diagonal matrix is also a diagonal matrix, and so this final expression is the product of three upper triangular matrices, and so is again upper triangular (Theorem PTMT [1415]). Thus the desired upper triangular matrix, T, is the matrix representation of S relative to the orthonormal basis D, $M_{D,D}^S$.

Subsection NM Normal Matrices

Normal matrices comprise a broad class of interesting matrices, many of which we have met already. But they are most interesting since they define exactly which matrices we can diagonalize via a unitary matrix. This is the upcoming Theorem OD [1429]. Here's the definition.

Definition NRML Normal Matrix

The square matrix A is normal if $A^*A = AA^*$.

So a normal matrix commutes with its adjoint. Part of the beauty of this definition is that it includes many other types of matrices. A diagonal matrix will commute with its adjoint, since the adjoint is again diagonal and the entries are just conjugates of the entries of the original diagonal matrix. A Hermitian (self-adjoint) matrix (Definition HM [486]) will trivially commute with its adjoint, since the two matrices are the same. A real, symmetric matrix is Hermitian, so these matrices are also normal. A unitary matrix (Definition UM [544]) has its adjoint as its inverse, and inverses commute (Theorem OSIS [541]), so unitary matrices are normal. Another class of normal matrices is the skew-symmetric matrices. However, these broad descriptions still do not capture all of the normal matrices, as the next example shows.

 \wedge

Example ANM A normal matrix Let

$$A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

Then

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

so we see by Definition NRML [1427] that A is normal. However, A is not symmetric (hence, as a real matrix, not Hermitian), not unitary, and not skew-symmetric.

Subsection OD Orthonormal Diagonalization

A diagonal matrix is very easy to work with in matrix multiplication (Example HPDM [1051]) and an orthonormal basis also has many advantages (Theorem COB [793]). How about converting a matrix to a diagonal matrix through a similarity transformation using a unitary matrix (i.e. build a diagonal matrix representation with an orthonormal matrix)?

That'd be fantastic! When can we do this? We can always accomplish this feat when the matrix is normal, and normal matrices are the only ones that behave this way. Here's the theorem.

Theorem OD Orthonormal Diagonalization

Suppose that A is a square matrix. Then there is a unitary matrix U and a diagonal matrix D, with diagonal entries equal to the eigenvalues of A, such that $U^*AU = D$ if and only if A is a normal matrix.

Proof (\Rightarrow) Suppose there is a unitary matrix U that diagonalizes A, resulting in D, i.e. $U^*AU = D$. We check the normality of A,

$$A^*A = I_n A^* I_n A I_n$$
Theorem MMIM [475] $= UU^* A^* UU^* A UU^*$ Definition UM [544] $= UU^* A^* UDU^*$ Theorem AA [443] $= U (U^* A U)^* DU^*$ Adjoint of a product $= U D^* DU^*$ Definition A [442] $= U \overline{D} DU^*$ Diagonal matrix $= U D \overline{D} U^*$ Property CMCN [1597]

 $= UD\left(\overline{D}\right)^{t}U^{*}$ $= UDD^*U^*$ $= UD \left(U^* A U \right)^* U^*$ $= UDU^*A^*(U^*)^*U^*$ $= UDU^*A^*UU^*$ $= UU^*AUU^*A^*UU^*$ $= I_n A I_n A^* I_n$ $= A A^*$

Diagonal matrix Definition A [442]

Adjoint of a product Theorem AA [443]

Definition UM [544] Theorem MMIM [475]

AA [443]

MMIM [475]

So by Definition NRML [1427], A is a normal matrix.

 (\Leftarrow) For the converse, suppose that A is a normal matrix. Whether or not A is normal, Theorem OBUTR [1424] provides a unitary matrix U and an upper triangular matrix T, whose diagonal entries are the eigenvalues of A, and such that $U^*AU = T$. With the added condition that A is normal, we will determine that the entries of T above the diagonal must be all zero. Here we go. First we show that T is normal.

$$T^{*}T = (U^{*}AU)^{*}U^{*}AU$$

$$= U^{*}A^{*}(U^{*})^{*}U^{*}AU$$
Adjoint of a product
$$= U^{*}A^{*}UU^{*}AU$$
Theorem AA [443]
$$= U^{*}A^{*}I_{n}AU$$
Definition UM [544]
$$= U^{*}A^{*}AU$$
Theorem MMIM [475]

 $= U^*AA^*U$ = U^*AI_nA^*U = U^*AUU^*A^*U = U^*AUU^*A^* (U^*)^* = U^*AU (U^*AU)^* = TT^* Definition NRML [1427] Theorem MMIM [475] Definition UM [544] Theorem AA [443] Adjoint of a product

So by Definition NRML [1427], T is a normal matrix.

We can translate the normality of T into the statement $TT^* - T^*T = \mathcal{O}$. We now establish an equality we will use repeatedly. For $1 \leq i \leq n$,

$$0 = [\mathcal{O}]_{ii}$$

= $[TT^* - T^*T]_{ii}$
= $[TT^*]_{ii} - [T^*T]_{ii}$
= $\sum_{k=1}^{n} [T]_{ik} [T^*]_{ki} - \sum_{k=1}^{n} [T^*]_{ik} [T]_{ki}$
= $\sum_{k=1}^{n} [T]_{ik} \overline{[T]_{ik}} - \sum_{k=1}^{n} \overline{[T]_{ki}} [T]_{ki}$
= $\sum_{k=i}^{n} [T]_{ik} \overline{[T]_{ik}} - \sum_{k=1}^{i} \overline{[T]_{ki}} [T]_{ki}$

Definition ZM [433] Definition NRML [1427] Definition MA [427]

Theorem EMP [471]

Definition A [442]

Definition UTM [1415]

$$= \sum_{k=i}^{n} |[T]_{ik}|^2 - \sum_{k=1}^{i} |[T]_{ki}|^2$$

Definition MCN [1601]

To conclude, we use the above equality repeatedly, beginning with i = 1, and discover, row by row, that the entries above the diagonal of T are all zero. The key observation is that a sum of squares can only equal zero when each term of the sum is zero. For i = 1 we have

$$0 = \sum_{k=1}^{n} |[T]_{1k}|^2 - \sum_{k=1}^{1} |[T]_{k1}|^2 = \sum_{k=2}^{n} |[T]_{1k}|^2$$

which forces the conclusions

$$[T]_{12} = 0$$
 $[T]_{13} = 0$ $[T]_{14} = 0$... $[T]_{1n} = 0$

For i = 2 we use the same equality, but also incorporate the portion of the above conclusions that says $[T]_{12} = 0$,

$$0 = \sum_{k=2}^{n} |[T]_{2k}|^2 - \sum_{k=1}^{2} |[T]_{k2}|^2 = \sum_{k=2}^{n} |[T]_{2k}|^2 - \sum_{k=2}^{2} |[T]_{k2}|^2 = \sum_{k=3}^{n} |[T]_{2k}|^2$$

which forces the conclusions

$$[T]_{23} = 0$$
 $[T]_{24} = 0$ $[T]_{25} = 0$... $[T]_{2n} = 0$

We can repeat this process for the subsequent values of i = 3, 4, 5..., n-1. Notice that it is critical we do this in order, since we need to employ portions of each of the previous conclusions about rows having zero entries in order to successfully get the same conclusion for later rows. Eventually, we conclude that all of the nondiagonal entries of T are zero, so the extra assumption of normality forces T to be diagonal.

We can rearrange the conclusion of this theorem to read $A = UDU^*$. Recall that a unitary matrix can be viewed as a geometry-preserving transformation (isometry), or more loosely as a rotation of sorts. Then a matrix-vector product, $A\mathbf{x}$, can be viewed instead as a sequence of three transformations. U^* is unitary, so is a rotation. Since D is diagonal, it just multiplies each entry of a vector by a scalar. Diagonal entries that are positive or negative, with absolute values bigger or smaller than 1 evoke descriptions like reflection, expansion and contraction. Generally we can say that D "stretches" a vector in each component. Final multiplication by U undoes (inverts) the rotation performed by U^* . So a normal matrix is a rotation-stretch-rotation transformation.

The orthonormal basis formed from the columns of U can be viewed as a system of mutually perpendicular axes. The rotation by U^* allows the transformation by A to be replaced by the simple transformation D along these axes, and then D brings the result back to the original coordinate system. For this reason Theorem OD [1429] is known as the Principal Axis Theorem.

The columns of the unitary matrix in Theorem OD [1429] create an especially nice basis for use with the normal matrix. We record this observation as a theorem.

Theorem OBNM

Orthonormal Bases and Normal Matrices

Suppose that A is a normal matrix of size n. Then there is an orthonormal basis of \mathbb{C}^n composed of eigenvectors of A.

Proof Let U be the unitary matrix promised by Theorem OD [1429] and let D be the resulting diagonal matrix. The desired set of vectors is formed by collecting the columns of U into a set. Theorem CUMOS [547] says this set of columns is orthonormal. Since U is nonsingular (Theorem UMI [546]), Theorem CNMB [789] says the set is a basis.

Since A is diagonalized by U, the diagonal entries of the matrix D are the eigenvalues of A. An argument exactly like the second half of the proof of Theorem DC [1039] shows that each vector of the basis is an eigenvector of A.

In a vague way Theorem OBNM [1434] is an improvement on Theorem HMOE [1022] which said that eigenvectors of a Hermitian matrix for different eigenvalues are always orthogonal. Hermitian matrices are normal and we see that we can find at least one basis where *every* pair of eigenvectors is orthogonal. Notice that this is not a generalization, since Theorem HMOE [1022] states a weak result which applies to many (but not all) pairs of eigenvectors, while Theorem OBNM [1434] is a seemingly stronger result, but only asserts that there is one collection of eigenvectors with the stronger property.

Section NLT Nilpotent Linear Transformations

This section is in draft form Nearly complete

We have seen that some matrices are diagonalizable and some are not. Some authors refer to a non-diagonalizable matrix as **defective**, but we will study them carefully anyway. Examples of such matrices include Example EMMS4 [969], Example HMEM5 [973], and Example CEMS6 [974]. Each of these matrices has at least one eigenvalue with geometric multiplicity strictly less than its algebraic multiplicity, and therefore Theorem DMFE [1044] tells us these matrices are not diagonalizable.

Given a square matrix A, it is likely similar to many, many other matrices. Of all these possibilities, which is the best? "Best" is a subjective term, but we might agree that a diagonal matrix is certainly a very nice choice. Unfortunately, as we have seen, this will not always be possible. What form of a matrix is "next-best"? Our goal, which will take us several sections to reach, is to show that every matrix is similar to a matrix that is "nearly-diagonal" (Section JCF [1517]). More precisely, every matrix is similar to a matrix with elements on the diagonal, and zeros and ones on the diagonal just above the main diagonal (the "super diagonal"), with zeros everywhere else. In the language of equivalence relations (see Theorem SER [1033]), we are determining a systematic representative for each equivalence class. Such a representative for a set of similar matrices is called a **canonical** form.

We have just discussed the determination of a canonical form as a question about matrices. However, we know that every square matrix creates a natural linear transformation (Theorem MBLT [1086]) and every linear transformation with identical domain and codomain has a square matrix representation for each choice of a basis, with a change of basis creating a similarity transformation (Theorem SCB [1377]). So we will state, and prove, theorems using the language of linear transformations on abstract vector spaces, while most of our examples will work with square matrices. You can, and should, mentally translate between the two settings frequently and easily.

Subsection NLT Nilpotent Linear Transformations

We will discover that nilpotent linear transformations are the essential obstacle in a nondiagonalizable linear transformation. So we will study them carefully first, both as an object of inherent mathematical interest, but also as the object at the heart of the argument that leads to a pleasing canonical form for any linear transformation. Once we understand these linear transformations thoroughly, we will be able to easily analyze the structure of any linear transformation.

Definition NLT Nilpotent Linear Transformation

Suppose that $T: V \to V$ is a linear transformation such that there is an integer p > 0 such that $T^p(\mathbf{v}) = \mathbf{0}$ for every $\mathbf{v} \in V$. The smallest p for which this condition is met is called the **index** of T.

Of course, the linear transformation T defined by $T(\mathbf{v}) = \mathbf{0}$ will qualify as nilpotent of index 1. But are there others?

Example NM64

Nilpotent matrix, size 6, index 4

Recall that our definitions and theorems are being stated for linear transformations on abstract vector spaces, while our examples will work with square matrices (and use the same terms interchangeably). In this case, to demonstrate the existence of nontrivial nilpotent linear transformations, we desire a matrix such that some power of the matrix is the zero matrix. Consider

$$A = \begin{bmatrix} -3 & 3 & -2 & 5 & 0 & -5 \\ -3 & 5 & -3 & 4 & 3 & -9 \\ -3 & 4 & -2 & 6 & -4 & -3 \\ -3 & 3 & -2 & 5 & 0 & -5 \\ -3 & 3 & -2 & 4 & 2 & -6 \\ -2 & 3 & -2 & 2 & 4 & -7 \end{bmatrix}$$

and compute powers of A,

Thus we can say that A is nilpotent of index 4.

Because it will presage some upcoming theorems, we will record some extra information about the eigenvalues and eigenvectors of A here. A has just one eigenvalue, $\lambda = 0$, with algebraic multiplicity 6 and geometric multiplicity 2. The eigenspace for this eigenvalue is

$$\mathcal{E}_{A}(0) = \left\langle \begin{bmatrix} 2\\2\\5\\2\\1\\0 \end{bmatrix}, \begin{bmatrix} -1\\-1\\-5\\-1\\0\\1 \end{bmatrix} \right\rangle$$

If there were degrees of singularity, we might say this matrix was *very* singular, since zero is an eigenvalue with maximum algebraic multiplicity (Theorem SMZE [1003], Theorem ME [1016]). Notice too that A is "far" from being diagonalizable (Theorem DMFE [1044]). \boxtimes

Another example.

Example NM62 Nilpotent matrix, size 6, index 2

Consider the matrix

$$B = \begin{bmatrix} -1 & 1 & -1 & 4 & -3 & -1 \\ 1 & 1 & -1 & 2 & -3 & -1 \\ -9 & 10 & -5 & 9 & 5 & -15 \\ -1 & 1 & -1 & 4 & -3 & -1 \\ 1 & -1 & 0 & 2 & -4 & 2 \\ 4 & -3 & 1 & -1 & -5 & 5 \end{bmatrix}$$

and compute the second power of B,

So B is nilpotent of index 2. Again, the only eigenvalue of B is zero, with algebraic multiplicity 6. The geometric multiplicity of the eigenvalue is 3, as seen in the eigenspace,

$$\mathcal{E}_{B}(0) = \left\langle \begin{bmatrix} 1\\3\\6\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\-4\\-7\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\2\\1\\0\\1\\0 \end{bmatrix} \right\rangle$$

Again, Theorem DMFE [1044] tells us that B is far from being diagonalizable.

On a first encounter with the definition of a nilpotent matrix, you might wonder if such a thing was possible at all. That a high power of a nonzero object could be zero is so very

 \boxtimes

different from our experience with scalars that it seems very unnatural. Hopefully the two previous examples were somewhat surprising. But we have seen that matrix algebra does not always behave the way we expect (Example MMNC [470]), and we also now recognize matrix products not just as arithmetic, but as function composition (Theorem MRCLT [1302]). We will now turn to some examples of nilpotent matrices which might be more transparent.

Definition JB Jordan Block

Given the scalar $\lambda \in \mathbb{C}$, the Jordan block $J_n(\lambda)$ is the $n \times n$ matrix defined by

$$\left[J_n\left(\lambda\right)\right]_{ij} = \begin{cases} \lambda & i = j\\ 1 & j = i+1\\ 0 & \text{otherwise} \end{cases}$$

(This definition contains Notation JB.)

Example JB4 Jordan block, size 4 A simple example of a Jordan block,

$$J_4(5) = \begin{bmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}$$

 \wedge

 \boxtimes

We will return to general Jordan blocks later, but in this section we are just interested in Jordan blocks where $\lambda = 0$. Here's an example of why we are specializing in these matrices now.

Example NJB5 Nilpotent Jordan block, size 5 Consider

$$J_5(0) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and compute powers,

So $J_5(0)$ is nilpotent of index 5. As before, we record some information about the eigenvalues and eigenvectors of this matrix. The only eigenvalue is zero, with algebraic multiplicity 5, the maximum possible (Theorem ME [1016]). The geometric multiplicity of this eigenvalue is just 1, the minimum possible (Theorem ME [1016]), as seen in the eigenspace,

$$\mathcal{E}_{J_{5}(0)}\left(0\right) = \left\langle \begin{bmatrix} 1\\0\\0\\0\\0\end{bmatrix} \right\rangle$$

There should not be any real surprises in this example. We can watch the ones in the powers of $J_5(0)$ slowly march off to the upper-right hand corner of the powers. In some vague way, the eigenvalues and eigenvectors of this matrix are equally extreme.

We can form combinations of Jordan blocks to build a variety of nilpotent matrices. Simply place Jordan blocks on the diagonal of a matrix with zeros everywhere else, to create a **block diagonal** matrix.

Example NM83 Nilpotent matrix, size 8, index 3 Consider the matrix

and compute powers,

So C is nilpotent of index 3. You should notice how block diagonal matrices behave in products (much like diagonal matrices) and that it was the largest Jordan block that determined the index of this combination. All eight eigenvalues are zero, and each of the three Jordan blocks contributes one eigenvector to a basis for the eigenspace, resulting in zero having a geometric multiplicity of 3.

It would appear that nilpotent matrices only have zero as an eigenvalue, so the algebraic multiplicity will be the maximum possible. However, by creating block diagonal matrices with Jordan blocks on the diagonal you should be able to attain any desired geometric multiplicity for this lone eigenvalue. Likewise, the size of the largest Jordan block employed will determine the index of the matrix. So nilpotent matrices with various combinations of index and geometric multiplicities are easy to manufacture. The predictable properties of block diagonal matrices in matrix products and eigenvector computations, along with the next theorem, make this possible. You might find Example NJB5 [1442] a useful companion to this proof.

Theorem NJB Nilpotent Jordan Blocks

The Jordan block $J_n(0)$ is nilpotent of index n.

Proof While not phrased as an if-then statement, the statement in the theorem is understood to mean that if we have a specific matrix $(J_n(0))$ then we need to establish it is nilpotent of a specified index. The first column of $J_n(0)$ is the zero vector, and the remaining n-1 columns are the standard unit vectors \mathbf{e}_i , $1 \le i \le n-1$ (Definition SUV [409]), which are also the first n-1 columns of the size n identity matrix I_n . As shorthand, write $J = J_n(0)$.

$$J = [\mathbf{0} | \mathbf{e}_1 | \mathbf{e}_2 | \mathbf{e}_3 | \dots | \mathbf{e}_{n-1}]$$

We will use the definition of matrix multiplication (Definition MM [468]), together with a proof by induction (Technique I [1630]), to study the powers of J. Our claim is that

$$J^{k} = [\mathbf{0} | \mathbf{0} | \dots | \mathbf{0} | \mathbf{e}_{1} | \mathbf{e}_{2} | \dots | \mathbf{e}_{n-k}]$$

for $1 \leq k \leq n$. For the base case, k = 1, and the definition of $J^1 = J_n(0)$ establishes the claim. For the induction step, first note that $J\mathbf{e}_1 = \mathbf{0}$ and $J\mathbf{e}_i = \mathbf{e}_{i-1}$ for $2 \leq i \leq n$. Then, assuming the claim is true for k, we examine the k + 1 case,

$$J^{k+1} = JJ^{k}$$

= $J [\mathbf{0} | \mathbf{0} | \dots | \mathbf{0} | \mathbf{e}_{1} | \mathbf{e}_{2} | \dots | \mathbf{e}_{n-k}]$ Induction Hypothesis
= $[J\mathbf{0} | J\mathbf{0} | \dots | J\mathbf{0} | J\mathbf{e}_{1} | J\mathbf{e}_{2} | \dots | J\mathbf{e}_{n-k}]$ Definition MM [468]

$$= \begin{bmatrix} \mathbf{0} | \mathbf{0} | \dots | \mathbf{0} | \mathbf{e}_{1} | \mathbf{e}_{2} | \dots | \mathbf{e}_{n-k-1} \end{bmatrix}$$
 Definition MVP [461]
$$= \begin{bmatrix} \mathbf{0} | \mathbf{0} | \dots | \mathbf{0} | \mathbf{e}_{1} | \mathbf{e}_{2} | \dots | \mathbf{e}_{n-(k+1)} \end{bmatrix}$$

This concludes the induction. So J^k has a nonzero entry (a one) in row n - k and column n, for $1 \le k \le n - 1$, and is therefore a nonzero matrix. However, $J^n = [\mathbf{0} | \mathbf{0} | \dots | \mathbf{0}] = \mathcal{O}$. By Definition NLT [1437], J is nilpotent of index n.

Subsection PNLT Properties of Nilpotent Linear Transformations

In this subsection we collect some basic properties of nilpotent linear transformations. After studying the examples in the previous section, some of these will be no surprise.

Theorem ENLT

Eigenvalues of Nilpotent Linear Transformations

Suppose that $T: V \to V$ is a nilpotent linear transformation and λ is an eigenvalue of T. Then $\lambda = 0$.

Proof Let **x** be an eigenvector of T for the eigenvalue λ , and suppose that T is nilpotent with index p. Then

 $\mathbf{0} = T^{p}\left(\mathbf{x}\right) \qquad \qquad \text{Definition NLT [1437]}$

$=\lambda^{p}\mathbf{x}$ Theorem EOMP [1006]

Because **x** is an eigenvector, it is nonzero, and therefore Theorem SMEZV [679] tells us that $\lambda^p = 0$ and so $\lambda = 0$.

Paraphrasing, all of the eigenvalues of a nilpotent linear transformation are zero. So in particular, the characteristic polynomial of a nilpotent linear transformation, T, on a vector space of dimension n, is simply $p_T(x) = x^n$.

The next theorem is not critical for what follows, but it will explain our interest in nilpotent linear transformations. More specifically, it is the first step in backing up the assertion that nilpotent linear transformations are the essential obstacle in a non-diagonalizable linear transformation. While it is not obvious from the statement of the theorem, it says that a nilpotent linear transformation is not diagonalizable, unless it is trivially so.

Theorem DNLT

Diagonalizable Nilpotent Linear Transformations

Suppose the linear transformation $T: V \to V$ is nilpotent. Then T is diagonalizable if and only if T is the zero linear transformation.

Proof We start with the easy direction. Let $n = \dim(V)$.

(\Leftarrow) The linear transformation $Z: V \to V$ defined by $Z(\mathbf{v}) = \mathbf{0}$ for all $\mathbf{v} \in V$ is nilpotent of index p = 1 and a matrix representation relative to any basis of V is the $n \times n$ zero matrix, \mathcal{O} . Quite obviously, the zero matrix is a diagonal matrix (Definition DIM [1037]) and hence Z is diagonalizable (Definition DZM [1037]).

(\Rightarrow) Assume now that T is diagonalizable, so $\gamma_T(\lambda) = \alpha_T(\lambda)$ for every eigenvalue λ (Theorem DMFE [1044]). By Theorem ENLT [1448], T has only one eigenvalue (zero),

which therefore must have algebraic multiplicity n (Theorem NEM [1015]). So the geometric multiplicity of zero will be n as well, $\gamma_T(0) = n$.

Let B be a basis for the eigenspace $\mathcal{E}_T(0)$. Then B is a linearly independent subset of V of size n, and by Theorem G [852] will be a basis for V. For any $\mathbf{x} \in B$ we have

$T\left(\mathbf{x}\right) = 0\mathbf{x}$	Definition EM $[963]$
= 0	Theorem ZSSM [676]

So T is identically zero on a basis for B, and since the action of a linear transformation on a basis determines all of the values of the linear transformation (Theorem LTDB [1094]), it is easy to see that $T(\mathbf{v}) = \mathbf{0}$ for every $\mathbf{v} \in V$.

So, other than one trivial case (the zero matrix), every nilpotent linear transformation is not diagonalizable. It remains to see what is so "essential" about this broad class of non-diagonalizable linear transformations. For this we now turn to a discussion of kernels of powers of nilpotent linear transformations, beginning with a result about general linear transformations that may not necessarily be nilpotent.

Theorem KPLT

Kernels of Powers of Linear Transformations

Suppose $T: V \to V$ is a linear transformation, where dim (V) = n. Then there is an integer $m, 0 \le m \le n$, such that

$$\{\mathbf{0}\} = \mathcal{K}(T^0) \subsetneq \mathcal{K}(T^1) \subsetneq \mathcal{K}(T^2) \subsetneq \cdots \subsetneq \mathcal{K}(T^m) = \mathcal{K}(T^{m+1}) = \mathcal{K}(T^{m+2}) = \cdots$$

Proof There are several items to verify in the conclusion as stated. First, we show that $\mathcal{K}(T^k) \subseteq \mathcal{K}(T^{k+1})$ for any k. Choose $\mathbf{z} \in \mathcal{K}(T^k)$. Then

$T^{k+1}\left(\mathbf{z}\right) = T\left(T^{k}\left(\mathbf{z}\right)\right)$	Definition LTC [1111]
$=T\left(0\right)$	Definition KLT [1138]
= 0	Theorem LTTZZ [1080]

So by Definition KLT [1138], $\mathbf{z} \in \mathcal{K}(T^{k+1})$ and by Definition SSET [1603] we have $\mathcal{K}(T^k) \subseteq \mathcal{K}(T^{k+1})$.

Second, we demonstrate the existence of a power m where consecutive powers result in equal kernels. A by-product will be the condition that m can be chosen so that $m \leq n$. To the contrary, suppose that

$$\{\mathbf{0}\} = \mathcal{K}(T^0) \subsetneq \mathcal{K}(T^1) \subsetneq \mathcal{K}(T^2) \subsetneq \cdots \subsetneq \mathcal{K}(T^{n-1}) \subsetneq \mathcal{K}(T^n) \subsetneq \mathcal{K}(T^{n+1}) \subsetneq \cdots$$

Since $\mathcal{K}(T^k) \subseteq \mathcal{K}(T^{k+1})$, Theorem PSSD [857] implies that dim $(\mathcal{K}(T^{k+1})) \ge \dim (\mathcal{K}(T^k)) + 1$. Repeated application of this observation yields

$$\dim \left(\mathcal{K}(T^{n+1}) \right) \ge \dim \left(\mathcal{K}(T^n) \right) + 1$$
$$\ge \dim \left(\mathcal{K}(T^{n-1}) \right) + 2$$

$$\geq \dim \left(\mathcal{K}(T^0) \right) + (n+1)$$
$$= \dim \left(\{ \mathbf{0} \} \right) + n + 1$$
$$= n+1$$

Thus, $\mathcal{K}(T^{n+1})$ has a basis of size at least n+1, which is a linearly independent set of size greater than n in the vector space V of dimension n. This contradicts Theorem G [852].

This contradiction yields the existence of an integer k such that $\mathcal{K}(T^k) = \mathcal{K}(T^{k+1})$, so we can define m to be smallest such integer with this property. From the argument above about dimensions resulting from a strictly increasing chain of subspaces, it should be clear that $m \leq n$.

It remains to show that once two consecutive kernels are equal, then all of the remaining kernels are equal. More formally, if $\mathcal{K}(T^m) = \mathcal{K}(T^{m+1})$, then $\mathcal{K}(T^m) = \mathcal{K}(T^{m+j})$ for all $j \geq 1$. We will give a proof by induction on j (Technique I [1630]). The base case (j = 1) is precisely our defining property for m.

In the induction step, we assume that $\mathcal{K}(T^m) = \mathcal{K}(T^{m+j})$ and endeavor to show that $\mathcal{K}(T^m) = \mathcal{K}(T^{m+j+1})$. At the outset of this proof we established that $\mathcal{K}(T^m) \subseteq \mathcal{K}(T^{m+j+1})$. So Definition SE [1604] requires only that we establish the subset inclusion in the opposite direction. To wit, choose $\mathbf{z} \in \mathcal{K}(T^{m+j+1})$. Then

$0=T^{m+j+1}\left(\mathbf{z}\right)$	Definition KLT [1138]
$=T^{m+j}\left(T\left(\mathbf{z}\right)\right)$	Definition LTC [1111]
$=T^{m}\left(T\left(\mathbf{z}\right) \right)$	Induction Hypothesis
$=T^{m+1}\left(\mathbf{z}\right)$	Definition LTC [1111]

$$=T^{m}\left(\mathbf{z}\right)$$
 Base Case

So by Definition KLT [1138], $\mathbf{z} \in \mathcal{K}(T^m)$ as desired.

We now specialize Theorem KPLT [1450] to the case of nilpotent linear transformations, which buys us just a bit more precision in the conclusion.

Theorem KPNLT

Kernels of Powers of Nilpotent Linear Transformations

Suppose $T: V \to V$ is a nilpotent linear transformation with index p and dim (V) = n. Then $0 \le p \le n$ and

$$\{\mathbf{0}\} = \mathcal{K}(T^0) \subsetneq \mathcal{K}(T^1) \subsetneq \mathcal{K}(T^2) \subsetneq \cdots \subsetneq \mathcal{K}(T^p) = \mathcal{K}(T^{p+1}) = \cdots = V$$

Proof Since $T^p = 0$ it follows that $T^{p+j} = 0$ for all $j \ge 0$ and thus $\mathcal{K}(T^{p+j}) = V$ for $j \ge 0$. So the value of m guaranteed by Theorem KPLT [1450] is at most p. The only remaining aspect of our conclusion that does not follow from Theorem KPLT [1450] is that m = p. To see this we must show that $\mathcal{K}(T^k) \subsetneq \mathcal{K}(T^{k+1})$ for $0 \le k \le p - 1$. If $\mathcal{K}(T^k) = \mathcal{K}(T^{k+1})$ for some k < p, then $\mathcal{K}(T^k) = \mathcal{K}(T^p) = V$. This implies that $T^k = 0$, violating the fact that T has index p. So the smallest value of m is indeed p, and we learn that p < n.

The structure of the kernels of powers of nilpotent linear transformations will be crucial to what follows. But immediately we can see a practical benefit. Suppose we are confronted with the question of whether or not an $n \times n$ matrix, A, is nilpotent or not. If we don't

 \square

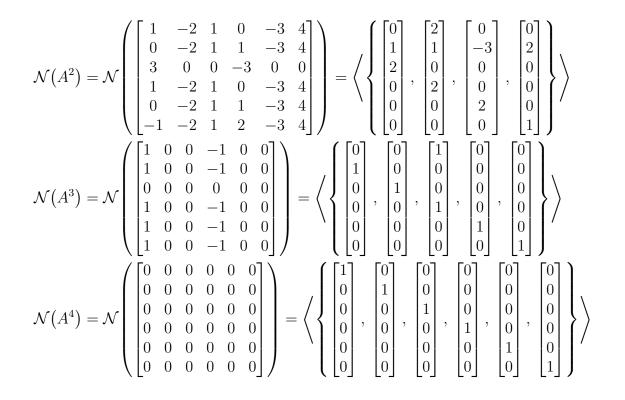
quickly find a low power that equals the zero matrix, when do we stop trying higher and higher powers? Theorem KPNLT [1453] gives us the answer: if we don't see a zero matrix by the time we finish computing A^n , then it is not going to ever happen. We'll now take a look at one example of Theorem KPNLT [1453] in action.

Example KPNLT

Kernels of powers of a nilpotent linear transformation

We will recycle the nilpotent matrix A of index 4 from Example NM64 [1437]. We now know that would have only needed to look at the first 6 powers of A if the matrix had not been nilpotent. We list bases for the null spaces of the powers of A. (Notice how we are using null spaces for matrices interchangeably with kernels of linear transformations, see Theorem KNSI [1311] for justification.)

$$\mathcal{N}(A) = \mathcal{N}\left(\begin{bmatrix} -3 & 3 & -2 & 5 & 0 & -5 \\ -3 & 5 & -3 & 4 & 3 & -9 \\ -3 & 4 & -2 & 6 & -4 & -3 \\ -3 & 3 & -2 & 5 & 0 & -5 \\ -3 & 3 & -2 & 4 & 2 & -6 \\ -2 & 3 & -2 & 2 & 4 & -7 \end{bmatrix} \right) = \left\langle \left\{ \begin{bmatrix} 2 \\ 2 \\ 5 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ -5 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\} \right\rangle$$



With the exception of some convenience scaling of the basis vectors in $\mathcal{N}(A^2)$ these are exactly the basis vectors described in Theorem BNS [334]. We can see that the dimension of $\mathcal{N}(A)$ equals the geometric multiplicity of the zero eigenvalue. Why is this not an accident? We can see the dimensions of the kernels consistently increasing, and we can see that $\mathcal{N}(A^4) = \mathbb{C}^6$. But Theorem KPNLT [1453] says a little more. Each successive kernel should be a superset of the previous one. We ought to be able to begin with a basis of $\mathcal{N}(A)$ and extend it to a basis of $\mathcal{N}(A^2)$. Then we should be able to extend a basis of $\mathcal{N}(A^2)$ into a basis of $\mathcal{N}(A^3)$, all with repeated applications of Theorem ELIS [850]. Verify the following,

$$\mathcal{N}(A) = \left\langle \left\{ \begin{bmatrix} 2\\2\\5\\2\\1\\0 \end{bmatrix}, \begin{bmatrix} -1\\-1\\-5\\-1\\0\\1 \end{bmatrix} \right\} \right\rangle$$
$$\mathcal{N}(A^{2}) = \left\langle \left\{ \begin{bmatrix} 2\\2\\5\\2\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} -1\\-1\\-5\\-1\\0\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} 0\\-3\\0\\0\\2\\0\\0\\1 \end{bmatrix}, \begin{bmatrix} 0\\2\\0\\0\\0\\1\\0\\1 \end{bmatrix} \right\} \right\rangle$$

Do not be concerned at the moment about how these bases were constructed since we are not describing the applications of Theorem ELIS [850] here. Do verify carefully for each alleged basis that, (1) it is a superset of the basis for the previous kernel, (2) the basis vectors really are members of the kernel of the right power of A, (3) the basis is a linearly independent set, (4) the size of the basis is equal to the size of the basis found previously for each kernel. With these verifications, Theorem G [852] will tell us that we have successfully demonstrated what Theorem KPNLT [1453] guarantees.

Subsection CFNLT Canonical Form for Nilpotent Linear Transformations

Our main purpose in this section is to find a basis so that a nilpotent linear transformation will have a pleasing, nearly-diagonal matrix representation. Of course, we will not have a definition for "pleasing," nor for "nearly-diagonal." But the short answer is that our preferred matrix representation will be built up from Jordan blocks, $J_n(0)$. Here's the theorem. You will find Example CFNLT [1468] helpful as you study this proof, since it uses the same notation, and is large enough to (barely) illustrate the full generality of the theorem.

Theorem CFNLT

Canonical Form for Nilpotent Linear Transformations

Suppose that $T: V \to V$ is a nilpotent linear transformation of index p. Then there is a basis for V so that the matrix representation, $M_{B,B}^T$, is block diagonal with each block being a Jordan block, $J_n(0)$. The size of the largest block is the index p, and the total number of blocks is the nullity of T, n(T).

Proof We will explicitly construct the desired basis, so the proof is constructive (Technique C [1618]), and can be used in practice. As we begin, the basis vectors will not be in the proper order, but we will rearrange them at the end of the proof. For convenience, define $n_i = n(T^i)$, so for example, $n_0 = 0$, $n_1 = n(T)$ and $n_p = n(T^p) = \dim(V)$. Define $s_i = n_i - n_{i-1}$, for $1 \le i \le p$, so we can think of s_i as "how much bigger" $\mathcal{K}(T^i)$ is than $\mathcal{K}(T^{i-1})$. In particular, Theorem KPNLT [1453] implies that $s_i > 0$ for $1 \le i \le p$.

We are going to build a set of vectors $\mathbf{z}_{i,j}$, $1 \leq i \leq p$, $1 \leq j \leq s_i$. Each $\mathbf{z}_{i,j}$ will be an element of $\mathcal{K}(T^i)$ and not an element of $\mathcal{K}(T^{i-1})$. In total, we will obtain a linearly independent set of $\sum_{i=1}^{p} s_i = \sum_{i=1}^{p} n_i - n_{i-1} = n_p - n_0 = \dim(V)$ vectors that form a basis of V. We construct this set in pieces, starting at the "wrong" end. Our procedure will build a series of subspaces, Z_i , each lying in between $\mathcal{K}(T^{i-1})$ and $\mathcal{K}(T^i)$, having bases $\mathbf{z}_{i,j}$, $1 \leq j \leq s_i$, and which together equal V as a direct sum. Now would be a good time to review the results on direct sums collected in Subsection PD.DS [863]. OK, here we go.

We build the subspace Z_p first (this is what we meant by "starting at the wrong end"). $\mathcal{K}(T^{p-1})$ is a proper subspace of $\mathcal{K}(T^p) = V$ (Theorem KPNLT [1453]). Theorem DSFOS [867] says that there is a subspace of V that will pair with the subspace $\mathcal{K}(T^{p-1})$ to form a direct sum of V. Call this subspace Z_p , and choose vectors $\mathbf{z}_{p,j}$, $1 \leq j \leq s_p$ as a basis of Z_p , which we will denote as B_p . Note that we have a fair amount of freedom in how to choose these first basis vectors. Several observations will be useful in the next step. First $V = \mathcal{K}(T^{p-1}) \oplus Z_p$. The basis $B_p = \{\mathbf{z}_{p,1}, \mathbf{z}_{p,2}, \mathbf{z}_{p,3}, \ldots, \mathbf{z}_{p,s_p}\}$ is linearly independent. For $1 \leq j \leq s_p, \mathbf{z}_{p,j} \in \mathcal{K}(T^p) = V$. Since the two subspaces of a direct sum have no nonzero vectors in common (Theorem DSZI [869]), for $1 \leq j \leq s_p, \mathbf{z}_{p,j} \notin \mathcal{K}(T^{p-1})$. That was comparably easy.

If obtaining Z_p was easy, getting Z_{p-1} will be harder. We will repeat the next step p-1 times, and so will do it carefully the first time. Eventually, Z_{p-1} will have dimension s_{p-1} . However, the first s_p vectors of a basis are straightforward. Define $\mathbf{z}_{p-1,j} = T(\mathbf{z}_{p,j})$, $1 \leq j \leq s_p$. Notice that we have no choice in creating these vectors, they are a consequence of our choices for $\mathbf{z}_{p,j}$. In retrospect (i.e. on a second reading of this proof), you will recognize this as the key step in realizing a matrix representation of a nilpotent linear transformation

with Jordan blocks. We need to know that this set of vectors in linearly independent, so start with a relation of linear dependence (Definition RLD [735]), and massage it,

$$\mathbf{0} = a_1 \mathbf{z}_{p-1,1} + a_2 \mathbf{z}_{p-1,2} + a_3 \mathbf{z}_{p-1,3} + \dots + a_{s_p} \mathbf{z}_{p-1,s_p}
= a_1 T (\mathbf{z}_{p,1}) + a_2 T (\mathbf{z}_{p,2}) + a_3 T (\mathbf{z}_{p,3}) + \dots + a_{s_p} T (\mathbf{z}_{p,s_p})
= T (a_1 \mathbf{z}_{p,1} + a_2 \mathbf{z}_{p,2} + a_3 \mathbf{z}_{p,3} + \dots + a_{s_p} \mathbf{z}_{p,s_p})$$
Theorem LTLC [1093]

Define $\mathbf{x} = a_1 \mathbf{z}_{p,1} + a_2 \mathbf{z}_{p,2} + a_3 \mathbf{z}_{p,3} + \cdots + a_{s_p} \mathbf{z}_{p,s_p}$. The statement just above means that $\mathbf{x} \in \mathcal{K}(T) \subseteq \mathcal{K}(T^{p-1})$ (Definition KLT [1138], Theorem KPNLT [1453]). As defined, \mathbf{x} is a linear combination of the basis vectors B_p , and therefore $\mathbf{x} \in Z_p$. Thus $\mathbf{x} \in \mathcal{K}(T^{p-1}) \cap Z_p$ (Definition SI [1608]). Because $V = \mathcal{K}(T^{p-1}) \oplus Z_p$, Theorem DSZI [869] tells us that $\mathbf{x} = \mathbf{0}$. Now we recognize the definition of \mathbf{x} as a relation of linear dependence on the linearly independent set B_p , and therefore $a_1 = a_2 = \cdots = a_{s_p} = 0$ (Definition LI [735]). This establishes the linear independence of $\mathbf{z}_{p-1,j}, 1 \leq j \leq s_p$ (Definition LI [735]).

We also need to know where the vectors $\mathbf{z}_{p-1,j}$, $1 \leq j \leq s_p$ live. First we demonstrate that they are members of $\mathcal{K}(T^{p-1})$.

$$T^{p-1} \left(\mathbf{z}_{p-1,j} \right) = T^{p-1} \left(T \left(\mathbf{z}_{p,j} \right) \right)$$
$$= T^{p} \left(\mathbf{z}_{p,j} \right)$$
$$= \mathbf{0}$$

So $\mathbf{z}_{p-1,j} \in \mathcal{K}(T^{p-1}), 1 \leq j \leq s_p$. However, we now show that these vectors are not elements

of $\mathcal{K}(T^{p-2})$. Suppose to the contrary (Technique CD [1623]) that $\mathbf{z}_{p-1,j} \in \mathcal{K}(T^{p-2})$. Then

$$\mathbf{0} = T^{p-2} \left(\mathbf{z}_{p-1,j} \right)$$
$$= T^{p-2} \left(T \left(\mathbf{z}_{p,j} \right) \right)$$
$$= T^{p-1} \left(\mathbf{z}_{p,j} \right)$$

which contradicts the earlier statement that $\mathbf{z}_{p,j} \notin \mathcal{K}(T^{p-1})$. So $\mathbf{z}_{p-1,j} \notin \mathcal{K}(T^{p-2}), 1 \leq j \leq s_p$.

Now choose a basis $C_{p-2} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_{n_{p-2}}\}$ for $\mathcal{K}(T^{p-2})$. We want to extend this basis by adding in the $\mathbf{z}_{p-1,j}$ to span a subspace of $\mathcal{K}(T^{p-1})$. But first we want to know that this set is linearly independent. Let a_k , $1 \leq k \leq n_{p-2}$ and b_j , $1 \leq j \leq s_p$ be the scalars in a relation of linear dependence,

$$\mathbf{0} = a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_{n_{p-2}} \mathbf{u}_{n_{p-2}} + b_1 \mathbf{z}_{p-1,1} + b_2 \mathbf{z}_{p-1,2} + \dots + b_{s_p} \mathbf{z}_{p-1,s_p}$$

Then,

$$\begin{aligned} \mathbf{0} &= T^{p-2} \left(\mathbf{0} \right) \\ &= T^{p-2} \left(a_1 \mathbf{u}_1 + a_2 \mathbf{u}_2 + \dots + a_{n_{p-2}} \mathbf{u}_{n_{p-2}} + b_1 \mathbf{z}_{p-1,1} + b_2 \mathbf{z}_{p-1,2} + \dots + b_{s_p} \mathbf{z}_{p-1,s_p} \right) \\ &= a_1 T^{p-2} \left(\mathbf{u}_1 \right) + a_2 T^{p-2} \left(\mathbf{u}_2 \right) + \dots + a_{n_{p-2}} T^{p-2} \left(\mathbf{u}_{n_{p-2}} \right) + \\ &\quad b_1 T^{p-2} \left(\mathbf{z}_{p-1,1} \right) + b_2 T^{p-2} \left(\mathbf{z}_{p-1,2} \right) + \dots + b_{s_p} T^{p-2} \left(\mathbf{z}_{p-1,s_p} \right) \\ &= a_1 \mathbf{0} + a_2 \mathbf{0} + \dots + a_{n_{p-2}} \mathbf{0} + b_1 T^{p-2} \left(\mathbf{z}_{p-1,1} \right) + b_2 T^{p-2} \left(\mathbf{z}_{p-1,2} \right) + \dots + b_{s_p} T^{p-2} \left(\mathbf{z}_{p-1,s_p} \right) \end{aligned}$$

$$= b_1 T^{p-2} (\mathbf{z}_{p-1,1}) + b_2 T^{p-2} (\mathbf{z}_{p-1,2}) + \dots + b_{s_p} T^{p-2} (\mathbf{z}_{p-1,s_p})$$

= $b_1 T^{p-2} (T (\mathbf{z}_{p,1})) + b_2 T^{p-2} (T (\mathbf{z}_{p,2})) + \dots + b_{s_p} T^{p-2} (T (\mathbf{z}_{p,s_p}))$
= $b_1 T^{p-1} (\mathbf{z}_{p,1}) + b_2 T^{p-1} (\mathbf{z}_{p,2}) + \dots + b_{s_p} T^{p-1} (\mathbf{z}_{p,s_p})$
= $T^{p-1} (b_1 \mathbf{z}_{p,1} + b_2 \mathbf{z}_{p,2} + \dots + b_{s_p} \mathbf{z}_{p,s_p})$

Define $\mathbf{y} = b_1 \mathbf{z}_{p,1} + b_2 \mathbf{z}_{p,2} + \cdots + b_{s_p} \mathbf{z}_{p,s_p}$. The statement just above means that $\mathbf{y} \in \mathcal{K}(T^{p-1})$ (Definition KLT [1138]). As defined, \mathbf{y} is a linear combination of the basis vectors B_p , and therefore $\mathbf{y} \in Z_p$. Thus $\mathbf{y} \in \mathcal{K}(T^{p-1}) \cap Z_p$. Because $V = \mathcal{K}(T^{p-1}) \oplus Z_p$, Theorem DSZI [869] tells us that $\mathbf{y} = \mathbf{0}$. Now we recognize the definition of \mathbf{y} as a relation of linear dependence on the linearly independent set B_p , and therefore $b_1 = b_2 = \cdots = b_{s_p} = 0$ (Definition LI [735]). Return to the full relation of linear dependence with both sets of scalars (the a_i and b_j). Now that we know that $b_j = 0$ for $1 \leq j \leq s_p$, this relation of linear dependence simplifies to a relation of linear dependence on just the basis C_{p-1} . Therefore, $a_i = 0, 1 \leq a_i \leq n_{p-1}$ and we have the desired linear independence.

Define a new subspace of $\mathcal{K}(T^{p-1})$ as

$$Q_{p-1} = \left\langle \left\{ \mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3, \, \dots, \, \mathbf{u}_{n_{p-1}}, \, \mathbf{z}_{p-1,1}, \, \mathbf{z}_{p-1,2}, \, \mathbf{z}_{p-1,3}, \, \dots, \, \mathbf{z}_{p-1,s_p} \right\} \right\rangle$$

By Theorem DSFOS [867] there exists a subspace of $\mathcal{K}(T^{p-1})$ which will pair with Q_{p-1} to form a direct sum. Call this subspace R_{p-1} , so by definition, $\mathcal{K}(T^{p-1}) = Q_{p-1} \oplus R_{p-1}$. We are interested in the dimension of R_{p-1} . Note first, that since the spanning set of Q_{p-1} is linearly independent, dim $(Q_{p-1}) = n_{p-2} + s_p$. Then

$$\dim (R_{p-1}) = \dim \left(\mathcal{K}(T^{p-1}) \right) - \dim (Q_{p-1}) \qquad \text{Theorem DSD [872]}$$

$$= n_{p-1} - (n_{p-2} + s_p)$$

= $(n_{p-1} - n_{p-2}) - s_p$
= $s_{p-1} - s_p$

Notice that if $s_{p-1} = s_p$, then R_{p-1} is trivial. Now choose a basis of R_{p-1} , and denote these $s_{p-1} - s_p$ vectors as \mathbf{z}_{p-1,s_p+1} , \mathbf{z}_{p-1,s_p+2} , \mathbf{z}_{p-1,s_p+3} , ..., $\mathbf{z}_{p-1,s_{p-1}}$. This is another occassion to notice that we have some freedom in this choice.

We now have $\mathcal{K}(T^{p-1}) = Q_{p-1} \oplus R_{p-1}$, and we have bases for each of the two subspaces. The union of these two bases will therefore be a linearly independent set in $\mathcal{K}(T^{p-1})$ with size

$$(n_{p-2} + s_p) + (s_{p-1} - s_p) = n_{p-2} + s_{p-1}$$

= $n_{p-2} + n_{p-1} - n_{p-2}$
= $n_{p-1} = \dim \left(\mathcal{K}(T^{p-1}) \right)$

So, by Theorem G [852], the following set is a basis of $\mathcal{K}(T^{p-1})$,

$$\{\mathbf{u}_1, \, \mathbf{u}_2, \, \mathbf{u}_3, \, \dots, \, \mathbf{u}_{n_{p-2}}, \, \mathbf{z}_{p-1,1}, \, \mathbf{z}_{p-1,2}, \, \dots, \, \mathbf{z}_{p-1,s_p}, \, \mathbf{z}_{p-1,s_p+1}, \, \mathbf{z}_{p-1,s_p+2}, \, \dots, \, \mathbf{z}_{p-1,s_{p-1}}\}$$

We built up this basis in three parts, we will now split it in half. Define the subspace Z_{p-1} by

$$Z_{p-1} = \langle B_{p-1} \rangle = \left\langle \left\{ \mathbf{z}_{p-1,1}, \, \mathbf{z}_{p-1,2}, \, \dots, \, \mathbf{z}_{p-1,s_{p-1}} \right\} \right\rangle$$

where we have implicitly denoted the basis as B_{p-1} . Then Theorem DSFB [865] allows us to split up the basis for $\mathcal{K}(T^{p-1})$ as $C_{p-1} \cup B_{p-1}$ and write

$$\mathcal{K}(T^{p-1}) = \mathcal{K}(T^{p-2}) \oplus Z_{p-1}$$

Whew! This is a good place to recap what we have achieved. The vectors $\mathbf{z}_{i,j}$ form bases for the subspaces Z_i and right now

$$V = \mathcal{K}(T^{p-1}) \oplus Z_p = \mathcal{K}(T^{p-2}) \oplus Z_{p-1} \oplus Z_p$$

The key feature of this decomposition of V is that the first s_p vectors in the basis for Z_{p-1} are outputs of the linear transformation T using the basis vectors of Z_p as inputs.

Now we want to further decompose $\mathcal{K}(T^{p-2})$ (into $\mathcal{K}(T^{p-3})$ and Z_{p-2}). The procedure is the same as above, so we will only sketch the key steps. Checking the details proceeds in the same manner as above. Technically, we could have set up the preceding as the induction step in a proof by induction (Technique I [1630]), but this probably would make the proof harder to understand.

Hit each element of B_{p-1} with T, to create vectors $\mathbf{z}_{p-2,j}$, $1 \leq j \leq s_{p-1}$. These vectors form a linearly independent set, and each is an element of $\mathcal{K}(T^{p-2})$, but not an element of $\mathcal{K}(T^{p-3})$. Grab a basis C_{p-3} of $\mathcal{K}(T^{p-3})$ and tack on the newly-created vectors $\mathbf{z}_{p-2,j}$, $1 \leq j \leq s_{p-1}$. This expanded set is linearly independent, and we can define a subspace Q_{p-2} using it as a basis. Theorem DSFOS [867] gives us a subspace R_{p-2} such that $\mathcal{K}(T^{p-2}) =$ $Q_{p-2} \oplus R_{p-2}$. Vectors $\mathbf{z}_{p-2,j}$, $s_{p-1} + 1 \leq j \leq s_{p-2}$ are chosen as a basis for R_{p-2} once the relevant dimensions have been verified. The union of C_{p-3} and $\mathbf{z}_{p-2,j}$, $1 \leq j \leq s_{p-2}$ then form a basis of $\mathcal{K}(T^{p-2})$, which can be split into two parts to yield the decomposition

$$\mathcal{K}(T^{p-2}) = \mathcal{K}(T^{p-3}) \oplus Z_{p-2}$$

Here Z_{p-2} is the subspace of $\mathcal{K}(T^{p-2})$ with basis $B_{p-2} = \{\mathbf{z}_{p-2,j} | 1 \le j \le s_{p-2}\}$. Finally,

$$V = \mathcal{K}(T^{p-1}) \oplus Z_p = \mathcal{K}(T^{p-2}) \oplus Z_{p-1} \oplus Z_p = \mathcal{K}(T^{p-3}) \oplus Z_{p-2} \oplus Z_{p-1} \oplus Z_p$$

Again, the key feature of this decomposition is that the first vectors in the basis of Z_{p-2} are outputs of T using vectors from the basis Z_{p-1} as inputs (and in turn, some of these inputs are outputs of T derived from inputs in Z_p).

Now assume we repeat this procedure until we decompose $\mathcal{K}(T^2)$ into subspaces $\mathcal{K}(T)$ and Z_2 . Finally, decompose $\mathcal{K}(T)$ into subspaces $\mathcal{K}(T^0) = \mathcal{K}(I_n) = \{\mathbf{0}\}$ and Z_1 , so that we recognize the vectors $\mathbf{z}_{1,j}$, $1 \leq j \leq s_1 = n_1$ as elements of $\mathcal{K}(T)$. The set

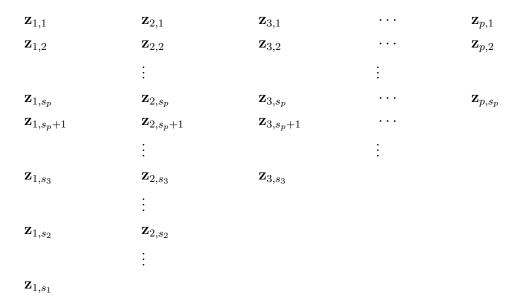
$$B = B_1 \cup B_2 \cup B_3 \cup \dots \cup B_p = \{ \mathbf{z}_{i,j} | 1 \le i \le p, \ 1 \le j \le s_i \}$$

is linearly independent by Theorem DSLI [870] and has size

$$\sum_{i=1}^{p} s_{i} = \sum_{i=1}^{p} n_{i} - n_{i-1} = n_{p} - n_{0} = \dim(V)$$

So by Theorem G [852], B is a basis of V. We desire a matrix representation of T relative to B (Definition MR [1286]), but first we will reorder the elements of B. The following display

lists the elements of B in the desired order, when read across the rows left-to-right in the usual way. Notice that we established the existence of these vectors column-by-column, and beginning on the right.



It is difficult to layout this table with the notation we have been using, but it would not be especially useful to invent some notation to overcome the difficulty. (One approach would be to define something like the inverse of the nonincreasing function, $i \to s_i$.) Do notice that there are $s_1 = n_1$ rows and p columns. Column i is the basis B_i . The vectors in the first column are elements of $\mathcal{K}(T)$. Each row is the same length, or shorter, than the one above it. If we apply T to any vector in the table, other than those in the first column, the output is the preceding vector in the row.

Now contemplate the matrix representation of T relative to B as we read across the rows of the table above. In the first row, $T(\mathbf{z}_{1,1}) = \mathbf{0}$, so the first column of the representation is the zero column. Next, $T(\mathbf{z}_{2,1}) = \mathbf{z}_{1,1}$, so the second column of the representation is a vector with a single one in the first entry, and zeros elsewhere. Next, $T(\mathbf{z}_{3,1}) = \mathbf{z}_{2,1}$, so column 3 of the representation is a zero, then a one, then all zeros. Continuing in this vein, we obtain the first p columns of the representation, which is the Jordan block $J_p(0)$ followed by rows of zeros.

When we apply T to the basis vectors of the second row, what happens? Applying T to the first vector, the result is the zero vector, so the representation gets a zero column. Applying T to the second vector in the row, the output is simply the first vector in that row, making the next column of the representation all zeros plus a lone one, sitting just above the diagonal. Continuing, we create a Jordan block, sitting on the diagonal of the matrix representation. It is not possible in general to state the size of this block, but since the second row is no longer than the first, it cannot have size larger than p.

Since there are as many rows as the dimension of $\mathcal{K}(T)$, the representation contains as many Jordan blocks as the nullity of T, n(T). Each successive block is smaller than the preceding one, with the first, and largest, having size p. The blocks are Jordan blocks since the basis vectors $\mathbf{z}_{i,j}$ were often defined as the result of applying T to other elements of the basis already determined, and then we rearranged the basis into an order that placed outputs of T just before their inputs, excepting the start of each row, which was an element of $\mathcal{K}(T)$.

The proof of Theorem CFNLT [1458] is constructive (Technique C [1618]), so we can use it to create bases of nilpotent linear transformations with pleasing matrix representations. Recall that Theorem DNLT [1449] told us that nilpotent linear transformations are almost never diagonalizable, so this is progress. As we have hinted before, with a nice representation of nilpotent matrices, it will not be difficult to build up representations of other non-diagonalizable matrices. Here is the promised example which illustrates the previous theorem. It is a useful companion to your study of the proof of Theorem CFNLT [1458].

Example CFNLT

Canonical form for a nilpotent linear transformation

The 6 × 6 matrix, A, of Example NM64 [1437] is nilpotent of index p = 4. If we define the linear transformation $T: \mathbb{C}^6 \to \mathbb{C}^6$ by $T(\mathbf{x}) = A\mathbf{x}$, then T is nilpotent of index 4 and we can seek a basis of \mathbb{C}^6 that yields a matrix representation with Jordan blocks on the diagonal. The nullity of T is 2, so from Theorem CFNLT [1458] we can expect the largest Jordan block to be $J_4(0)$, and there will be just two blocks. This only leaves enough room for the second block to have size 2.

We will recycle the bases for the null spaces of the powers of A from Example KPNLT [1454] rather than recomputing them here. We will also use the same notation used in the proof of Theorem CFNLT [1458].

To begin, $s_4 = n_4 - n_3 = 6 - 5 = 1$, so we need one vector of $\mathcal{K}(T^4) = \mathbb{C}^6$, that is not in $\mathcal{K}(T^3)$, to be a basis for Z_4 . We have a lot of latitude in this choice, and we have not

described any sure-fire method for constructing a vector *outside* of a subspace. Looking at the basis for $\mathcal{K}(T^3)$ we see that if a vector is in this subspace, and has a nonzero value in the first entry, then it must also have a nonzero value in the fourth entry. So the vector

$$\mathbf{z}_{4,1} = \begin{bmatrix} 1\\0\\0\\0\\0\\0\end{bmatrix}$$

will not be an element of $\mathcal{K}(T^3)$ (notice that many other choices could be made here, so our basis will not be unique). This completes the determination of $Z_p = Z_4$.

Next, $s_3 = n_3 - n_2 = 5 - 4 = 1$, so we again need just a single basis vector for Z_3 . We start by evaluating T with each basis vector of Z_4 ,

$$\mathbf{z}_{3,1} = T\left(\mathbf{z}_{4,1}\right) = A\mathbf{z}_{4,1} = \begin{bmatrix} -3\\ -3\\ -3\\ -3\\ -3\\ -2 \end{bmatrix}$$

Since $s_3 = s_4$, the subspace R_3 is trivial, and there is nothing left to do, $\mathbf{z}_{3,1}$ is the lone basis vector of Z_3 .

Now $s_2 = n_2 - n_1 = 4 - 2 = 2$, so the construction of Z_2 will not be as simple as the construction of Z_3 . We first apply T to the basis vector of Z_2 ,

$$\mathbf{z}_{2,1} = T\left(\mathbf{z}_{3,1}\right) = A\mathbf{z}_{3,1} = \begin{bmatrix} 1\\0\\3\\1\\0\\-1 \end{bmatrix}$$

The two basis vectors of $\mathcal{K}(T^1)$, together with $\mathbf{z}_{2,1}$, form a basis for Q_2 . Because dim $(\mathcal{K}(T^2))$ dim $(Q_2) = 4 - 3 = 1$ we need only find a single basis vector for R_2 . This vector must be an element of $\mathcal{K}(T^2)$, but not an element of Q_2 . Again, there is a variety of vectors that fit this description, and we have no precise algorithm for finding them. Since they are plentiful, they are not too hard to find. We add up the four basis vectors of $\mathcal{K}(T^2)$, ensuring an element of $\mathcal{K}(T^2)$. Then we check to see if the vector is a linear combination of three vectors: the two basis vectors of $\mathcal{K}(T^1)$ and $\mathbf{z}_{2,1}$. Having passed the tests, we have chosen

$$\mathbf{z}_{2,2} = \begin{vmatrix} 2\\1\\2\\2\\1\\1\end{vmatrix}$$

Thus, $Z_2 = \langle \{ \mathbf{z}_{2,1}, \, \mathbf{z}_{2,2} \} \rangle$.

Lastly, $s_1 = n_1 - n_0 = 2 - 0 = 2$. Since $s_2 = s_1$, we again have a trivial R_1 and need only complete our basis by evaluating the basis vectors of Z_2 with T,

$$\mathbf{z}_{1,1} = T(\mathbf{z}_{2,1}) = A\mathbf{z}_{2,1} = \begin{bmatrix} 1\\1\\0\\1\\1\\1 \end{bmatrix}$$
$$\mathbf{z}_{1,2} = T(\mathbf{z}_{2,2}) = A\mathbf{z}_{2,2} = \begin{bmatrix} -2\\-2\\-5\\-2\\-1\\0 \end{bmatrix}$$

Now we reorder these vectors as the desired basis,

$$B = \{\mathbf{z}_{1,1}, \, \mathbf{z}_{2,1}, \, \mathbf{z}_{3,1}, \, \mathbf{z}_{4,1}, \, \mathbf{z}_{1,2}, \, \mathbf{z}_{2,2}\}$$

We now apply Definition MR [1286] to build a matrix representation of T relative to B,

$$\rho_B \left(T \left(\mathbf{z}_{1,1} \right) \right) = \rho_B \left(A \mathbf{z}_{1,1} \right) = \rho_B \left(\mathbf{0} \right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{z}_{2,1} \right) \right) = \rho_B \left(A \mathbf{z}_{2,1} \right) = \rho_B \left(\mathbf{z}_{1,1} \right) = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{z}_{3,1} \right) \right) = \rho_B \left(A \mathbf{z}_{3,1} \right) = \rho_B \left(\mathbf{z}_{2,1} \right) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\rho_B \left(T \left(\mathbf{z}_{4,1} \right) \right) = \rho_B \left(A \mathbf{z}_{4,1} \right) = \rho_B \left(\mathbf{z}_{3,1} \right) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{z}_{1,2} \right) \right) = \rho_B \left(A \mathbf{z}_{1,2} \right) = \rho_B \left(\mathbf{0} \right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{z}_{2,2} \right) \right) = \rho_B \left(A \mathbf{z}_{2,2} \right) = \rho_B \left(\mathbf{z}_{1,2} \right) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Installing these vectors as the columns of the matrix representation we have

which is a block diagonal matrix with Jordan blocks $J_4(0)$ and $J_2(0)$. If we constructed the matrix S having the vectors of B as columns, then Theorem SCB [1377] tells us that a similarity transformation with S relates the original matrix representation of T with the matrix representation consisting of Jordan blocks., i.e. $S^{-1}AS = M_{B,B}^T$.

Notice that constructing interesting examples of matrix representations requires domains with dimensions bigger than just two or three. Going forward we will see several more big examples.

Section IS Invariant Subspaces

This section is in draft form Nearly complete

We have seen in Section NLT [1435] that nilpotent linear transformations are almost never diagonalizable (Theorem DNLT [1449]), yet have matrix representations that are very nearly diagonal (Theorem CFNLT [1458]). Our goal in this section, and the next (Section JCF [1517]), is to obtain a matrix representation of *any* linear transformation that is very nearly diagonal. A key step in reaching this goal is an understanding of invariant subspaces, and a particular type of invariant subspace that contains vectors known as "generalized eigenvectors."

Subsection IS Invariant Subspaces

As is often the case, we start with a definition.

Definition IS

Invariant Subspace

Suppose that $T: V \to V$ is a linear transformation and W is a subspace of V. Suppose further that $T(\mathbf{w}) \in W$ for every $\mathbf{w} \in W$. Then W is an **invariant subspace** of V relative to T.

We do not have any special notation for an invariant subspace, so it is important to recognize that an invariant subspace is always relative to both a superspace (V) and a linear transformation (T), which will sometimes not be mentioned, yet will be clear from the context. Note also that the linear transformation involved must have an equal domain and codomain — the definition would not make much sense if our outputs were not of the same type as our inputs.

As usual, we begin with an example that demonstrates the existence of invariant subspaces. We will return later to understand how this example was constructed, but for now, just understand how we check the existence of the invariant subspaces.

Example TIS Two invariant subspaces

Consider the linear transformation $T: \mathbb{C}^4 \to \mathbb{C}^4$ defined by $T(\mathbf{x}) = A\mathbf{x}$ where A is given by

$$A = \begin{bmatrix} -8 & 6 & -15 & 9\\ -8 & 14 & -10 & 18\\ 1 & 1 & 3 & 0\\ 3 & -8 & 2 & -11 \end{bmatrix}$$

Define (with zero motivation),

$$\mathbf{w_1} = \begin{bmatrix} -7\\ -2\\ 3\\ 0 \end{bmatrix} \qquad \qquad \mathbf{w_2} = \begin{bmatrix} -1\\ -2\\ 0\\ 1 \end{bmatrix}$$

and set $W = \langle \{\mathbf{w}_1, \mathbf{w}_2\} \rangle$. We verify that W is an invariant subspace of \mathbb{C}^4 with respect to T. By the definition of W, any vector chosen from W can be written as a linear combination of \mathbf{w}_1 and \mathbf{w}_2 . Suppose that $\mathbf{w} \in W$, and then check the details of the following verification,

$$T (\mathbf{w}) = T (a_1 \mathbf{w}_1 + a_2 \mathbf{w}_2)$$

Definition SS [709]
$$= a_1 T (\mathbf{w}_1) + a_2 T (\mathbf{w}_2)$$

Theorem LTLC [1093]
$$= a_1 \begin{bmatrix} -1 \\ -2 \\ 0 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 5 \\ -2 \\ -3 \\ 2 \end{bmatrix}$$

$$= a_1 \mathbf{w}_2 + a_2 ((-1) \mathbf{w}_1 + 2 \mathbf{w}_2)$$

$$= (-a_2) \mathbf{w}_1 + (a_1 + 2a_2) \mathbf{w}_2$$

 $\in W$
Definition SS [709]

So, by Definition IS [1476], W is an invariant subspace of \mathbb{C}^4 relative to T. In an entirely similar manner we construct another invariant subspace of T.

With zero motivation, define

$$\mathbf{x_1} = \begin{bmatrix} -3\\ -1\\ 1\\ 0 \end{bmatrix} \qquad \qquad \mathbf{x_2} = \begin{bmatrix} 0\\ -1\\ 0\\ 1 \end{bmatrix}$$

and set $X = \langle \{\mathbf{x}_1, \mathbf{x}_2\} \rangle$. We verify that X is an invariant subspace of \mathbb{C}^4 with respect to T. By the definition of X, any vector chosen from X can be written as a linear combination of \mathbf{x}_1 and \mathbf{x}_2 . Suppose that $\mathbf{x} \in X$, and then check the details of the following verification,

$$T (\mathbf{x}) = T (b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2)$$
 Definition SS [709]
= $b_1 T (\mathbf{x}_1) + b_2 T (\mathbf{x}_2)$ Theorem LTLC [1093
= $b_1 \begin{bmatrix} 3 \\ 0 \\ -1 \\ 1 \end{bmatrix} + b_2 \begin{bmatrix} 3 \\ 4 \\ -1 \\ -3 \end{bmatrix}$
= $b_1 ((-1)\mathbf{x}_1 + \mathbf{x}_2) + b_2 ((-1)\mathbf{x}_1 + (-3)\mathbf{x}_2)$
= $(-b_1 - b_2)\mathbf{x}_1 + (b_1 - 3b_2)\mathbf{x}_2$
 $\in X$ Definition SS [709]

So, by Definition IS [1476], X is an invariant subspace of \mathbb{C}^4 relative to T.

There is a bit of magic in each of these verifications where the two outputs of T happen to equal linear combinations of the two inputs. But this is the essential nature of an invariant subspace. We'll have a peek under the hood later, and it won't look so magical after all.

As a hint of things to come, verify that $B = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{x}_1, \mathbf{x}_2\}$ is a basis of \mathbb{C}^4 . Splitting this basis in half, Theorem DSFB [865], tells us that $\mathbb{C}^4 = W \oplus X$. To see why a decomposition of a vector space into a direct sum of invariant subspaces might be interesting, construct the matrix representation of T relative to $B, M_{B,B}^T$. Hmmmmm.

Example TIS [1476] is a bit mysterious at this stage. Do we know any other examples of invariant subspaces? Yes, as it turns out, we have already seen quite a few. We'll give some examples now, and in more general situations, describe broad classes of invariant subspaces with theorems. First up is eigenspaces.

Theorem EIS

Eigenspaces are Invariant Subspaces

Suppose that $T: V \to V$ is a linear transformation with eigenvalue λ and associated eigenspace $\mathcal{E}_T(\lambda)$. Let W be any subspace of $\mathcal{E}_T(\lambda)$. Then W is an invariant subspace of V relative to T.

Proof Choose $\mathbf{w} \in W$. Then

$T\left(\mathbf{w}\right) = \lambda \mathbf{w}$	Definition EELT [1356]
$\in W$	Property SC [660]

So by Definition IS [1476], W is an invariant subspace of V relative to T.

Theorem EIS [1479] is general enough to determine that an entire eigenspace is an in-

variant subspace, or that simply the span of a single eigenvector is an invariant subspace. It is not always the case that any subspace of an invariant subspace is again an invariant subspace, but eigenspaces do have this property. Here is an example of the theorem, which also allows us to very quickly build several several invariant (4x4, 2 evs, 1 2x2 jordan, 1 2x2 diag)

Example EIS

Eigenspaces as invariant subspaces

Define the linear transformation $S: M_{22} \to M_{22}$ by

$$S\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} -2a + 19b - 33c + 21d & -3a + 16b - 24c + 15d \\ -2a + 9b - 13c + 9d & -a + 4b - 6c + 5d \end{bmatrix}$$

Build a matrix representation of S relative to the standard basis (Definition MR [1286], Example BM [780]) and compute eigenvalues and eigenspaces of S with the computational techniques of Chapter E [944] in concert with Theorem EER [1384]. Then

$$\mathcal{E}_{S}(1) = \left\langle \left\{ \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \right\} \right\rangle \qquad \qquad \mathcal{E}_{S}(2) = \left\langle \left\{ \begin{bmatrix} 6 & 3 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -9 & -3 \\ 0 & 1 \end{bmatrix} \right\} \right\rangle$$

So by Theorem EIS [1479], both $\mathcal{E}_{S}(1)$ and $\mathcal{E}_{S}(2)$ are invariant subspaces of M_{22} relative to S. However, Theorem EIS [1479] provides even more invariant subspaces. Since $\mathcal{E}_{S}(1)$ has dimension 1, it has no interesting subspaces, however $\mathcal{E}_{S}(2)$ has dimension 2 and has a plethora of subspaces. For example, set

$$\mathbf{u} = 2 \begin{bmatrix} 6 & 3 \\ 1 & 0 \end{bmatrix} + 3 \begin{bmatrix} -9 & -3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -6 & -3 \\ 2 & 3 \end{bmatrix}$$

and define $U = \langle \{\mathbf{u}\} \rangle$. Then since U is a subspace of $\mathcal{E}_S(2)$, Theorem EIS [1479] says that U is an invariant subspace of M_{22} (or we could check this claim directly based simply on the fact that **u** is an eigenvector of S).

For every linear transformation there are some obvious, trivial invariant subspaces. Suppose that $T: V \to V$ is a linear transformation. Then simply because T is a function (Definition LT [1072]), the subspace V is an invariant subspace of T. In only a minor twist on this theme, the range of T, $\mathcal{R}(T)$, is an invariant subspace of T by Definition RLT [1178]. Finally, Theorem LTTZZ [1080] provides the justification for claiming that $\{0\}$ is an invariant subspace of T.

That the trivial subspace is always an invariant subspace is a special case of the next theorem. As an easy exercise before reading the next theorem, prove that the kernel of a linear transformation (Definition KLT [1138]), $\mathcal{K}(T)$, is an invariant subspace. We'll wait.

Theorem KPIS

Kernels of Powers are Invariant Subspaces

Suppose that $T: V \to V$ is a linear transformation. Then $\mathcal{K}(T^k)$ is an invariant subspace of V.

Proof Suppose that $\mathbf{z} \in \mathcal{K}(T^k)$. Then

$$T^{k}(T(\mathbf{z})) = T^{k+1}(\mathbf{z}) \qquad \text{Definition LTC [1111]} \\ = T(T^{k}(\mathbf{z})) \qquad \text{Definition LTC [1111]} \\ = T(\mathbf{0}) \qquad \text{Definition KLT [1138]}$$

= 0

Theorem LTTZZ [1080]

So by Definition KLT [1138], we see that $T(\mathbf{z}) \in \mathcal{K}(T^k)$. Thus $\mathcal{K}(T^k)$ is an invariant subspace of V relative to T (Definition IS [1476]).

Two interesting special cases of Theorem KPIS [1481] occur when choose k = 0 and k = 1. Rather than give an example of this theorem, we will refer you back to Example KPNLT [1454] where we work with null spaces of the first four powers of a nilpotent matrix. By Theorem KPIS [1481] each of these null spaces is an invariant subspace of the associated linear transformation.

Here's one more example of invariant subspaces we have encountered previously.

Example ISJB

Invariant subspaces and Jordan blocks

Refer back to Example CFNLT [1468]. We decomposed the vector space \mathbb{C}^6 into a direct sum of the subspaces Z_1 , Z_2 , Z_3 , Z_4 . The union of the basis vectors for these subspaces is a basis of \mathbb{C}^6 , which we reordered prior to building a matrix representation of the linear transformation T. A principal reason for this reordering was to create invariant subspaces (though it was not obvious then). Define

$$X_{1} = \langle \{\mathbf{z}_{1,1}, \, \mathbf{z}_{2,1}, \, \mathbf{z}_{3,1}, \, \mathbf{z}_{4,1}\} \rangle = \left\langle \left\{ \begin{bmatrix} 1\\1\\0\\1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} 1\\0\\3\\1\\0\\-1 \end{bmatrix}, \begin{bmatrix} -3\\-3\\-3\\-3\\-3\\-2 \end{bmatrix}, \begin{bmatrix} 1\\0\\0\\0\\0\\0\\0\\0 \end{bmatrix} \right\} \right\rangle$$
$$X_{2} = \left\langle \{\mathbf{z}_{1,2}, \, \mathbf{z}_{2,2}\} \right\rangle = \left\langle \left\{ \begin{bmatrix} -2\\-2\\-5\\-2\\-5\\-2\\-1\\0 \end{bmatrix}, \begin{bmatrix} 2\\1\\2\\2\\1\\1 \end{bmatrix} \right\} \right\rangle$$

Recall from the proof of Theorem CFNLT [1458] or the computations in Example CFNLT [1468] that first elements of X_1 and X_2 are in the kernel of T, $\mathcal{K}(T)$, and each element of X_1 and X_2 is the output of T when evaluated with the subsequent element of the set. This was by design, and it is this feature of these basis vectors that leads to the nearly diagonal matrix representation with Jordan blocks. However, we also recognize now that this property of these basis vectors allow us to conclude easily that X_1 and X_2 are invariant subspaces of \mathbb{C}^6 relative to T.

Furthermore, $\mathbb{C}^6 = X_1 \oplus X_2$ (Theorem DSFB [865]). So the domain of T is the direct sum of invariant subspaces and the resulting matrix representation has a block diagonal form.

Hmmmm.

\boxtimes

Subsection GEE Generalized Eigenvectors and Eigenspaces

We now define a new type of invariant subspace and explore its key properties. This generalization of eigenvalues and eigenspaces will allow us to move from diagonal matrix representations of diagonalizable matrices to nearly diagonal matrix representations of arbitrary matrices. Here are the definitions.

Definition GEV Generalized Eigenvector

Suppose that $T: V \to V$ is a linear transformation. Suppose further that for $\mathbf{x} \neq \mathbf{0}$, $(T - \lambda I_V)^k(\mathbf{x}) = \mathbf{0}$ for some k > 0. Then \mathbf{x} is a **generalized eigenvector** of T with eigenvalue λ .

Definition GES

Generalized Eigenspace

Suppose that $T: V \to V$ is a linear transformation. Define the **generalized eigenspace** of T for λ as

$$\mathcal{G}_{T}(\lambda) = \left\{ \mathbf{x} | \left(T - \lambda I_{V}\right)^{k}(\mathbf{x}) = \mathbf{0} \text{ for some } k \ge 0 \right\}$$

(This definition contains Notation GES.)

So the generalized eigenspace is composed of generalized eigenvectors, plus the zero vector. As the name implies, the generalized eigenspace is a subspace of V. But more topically, it is an invariant subspace of V relative to T.

Theorem GESIS

Generalized Eigenspace is an Invariant Subspace

Suppose that $T: V \to V$ is a linear transformation. Then the generalized eigenspace $\mathcal{G}_T(\lambda)$ is an invariant subspace of V relative to T.

Proof First we establish that $\mathcal{G}_T(\lambda)$ is a subspace of V. First $(T - \lambda I_V)^1(\mathbf{0}) = \mathbf{0}$ by Theorem LTTZZ [1080], so $\mathbf{0} \in \mathcal{G}_T(\lambda)$.

Suppose that $\mathbf{x}, \mathbf{y} \in \mathcal{G}_T(\lambda)$. Then there are integers k, ℓ such that $(T - \lambda I_V)^k (\mathbf{x}) = \mathbf{0}$ and $(T - \lambda I_V)^\ell (\mathbf{y}) = \mathbf{0}$. Set $m = k + \ell$,

$$(T - \lambda I_V)^m (\mathbf{x} + \mathbf{y}) = (T - \lambda I_V)^m (\mathbf{x}) + (T - \lambda I_V)^m (\mathbf{y}) \qquad \text{Definition LT [1072]} = (T - \lambda I_V)^{k+\ell} (\mathbf{x}) + (T - \lambda I_V)^{k+\ell} (\mathbf{y}) = (T - \lambda I_V)^\ell \left((T - \lambda I_V)^k (\mathbf{x}) \right) + (T - \lambda I_V)^k (\mathbf{y}) \qquad \text{Definition LTC [1111]} = (T - \lambda I_V)^\ell (\mathbf{0}) + (T - \lambda I_V)^k (\mathbf{0}) \qquad \text{Definition GES [1484]} = \mathbf{0} + \mathbf{0} \qquad \text{Theorem LTTZZ [1080]}$$

 \triangle

= 0

Property Z [660]

So $\mathbf{x} + \mathbf{y} \in \mathcal{G}_T(\lambda)$. Suppose that $\mathbf{x} \in \mathcal{G}_T(\lambda)$ and $\alpha \in \mathbb{C}$. Then there is an integer k such that $(T - \lambda I_V)^k (\mathbf{x}) = \mathbf{0}$.

$$(T - \lambda I_V)^k (\alpha \mathbf{x}) = \alpha (T - \lambda I_V)^k (\mathbf{x})$$
 Definition LT [1072]
= $\alpha \mathbf{0}$ Definition GES [1484]
= $\mathbf{0}$ Theorem ZVSM [677]

So $\alpha \mathbf{x} \in \mathcal{G}_T(\lambda)$. By Theorem TSS [698], $\mathcal{G}_T(\lambda)$ is a subspace of V.

Now we show that $\mathcal{G}_T(\lambda)$ is invariant relative to T. Suppose that $\mathbf{x} \in \mathcal{G}_T(\lambda)$. Then by Definition GES [1484] there is an integer k such that $(T - \lambda I_V)^k(\mathbf{x}) = \mathbf{0}$. The following argument is due to Zoltan Toth.

$$(T - \lambda I_V)^k (T (\mathbf{x})) = (T - \lambda I_V)^k (T (\mathbf{x})) - \mathbf{0} \qquad \text{Property Z [660]}$$

$$= (T - \lambda I_V)^k (T (\mathbf{x})) - \lambda \mathbf{0} \qquad \text{Theorem ZVSM [677]}$$

$$= (T - \lambda I_V)^k (T (\mathbf{x})) - \lambda (T - \lambda I_V)^k (\mathbf{x}) \qquad \text{Definition GES [1484]}$$

$$= (T - \lambda I_V)^k (T (\mathbf{x})) - (T - \lambda I_V)^k (\lambda \mathbf{x}) \qquad \text{Definition LT [1072]}$$

$$= (T - \lambda I_V)^k (T (\mathbf{x}) - \lambda \mathbf{x}) \qquad \text{Definition LT [1072]}$$

$$= (T - \lambda I_V)^k ((T - \lambda I_V) (\mathbf{x})) \qquad \text{Definition LTA [1105]}$$

$$= (T - \lambda I_V)^{k+1} (\mathbf{x}) \qquad \text{Definition LTC [1111]}$$

$$= (T - \lambda I_V) \left((T - \lambda I_V)^k (\mathbf{x}) \right)$$

Definition LTC [1111]
$$= (T - \lambda I_V) (\mathbf{0})$$

Definition GES [1484]
Theorem LTTZZ [1080]

This qualifies $T(\mathbf{x})$ for membership in $\mathcal{G}_T(\lambda)$, so by Definition GES [1484], $\mathcal{G}_T(\lambda)$ is invariant relative to T.

Before we compute some generalized eigenspaces, we state and prove one theorem that will make it much easier to create a generalized eigenspace, since it will allow us to use tools we already know well, and will remove some the ambiguity of the clause "for some k" in the definition.

Theorem GEK Generalized Eigenspace as a Kernel

Suppose that $T: V \to V$ is a linear transformation, dim (V) = n, and λ is an eigenvalue of T. Then $\mathcal{G}_T(\lambda) = \mathcal{K}((T - \lambda I_V)^n)$.

Proof The conclusion of this theorem is a set equality, so we will apply Definition SE [1604] by establishing two set inclusions. First, suppose that $\mathbf{x} \in \mathcal{G}_T(\lambda)$. Then there is an integer k such that $(T - \lambda I_V)^k (\mathbf{x}) = \mathbf{0}$. This is equivalent to the statement that $\mathbf{x} \in \mathcal{K}((T - \lambda I_V)^k)$. No matter what the value of k is, Theorem KPLT [1450] gives

$$\mathbf{x} \in \mathcal{K}\left(\left(T - \lambda I_V\right)^k\right) \subseteq \mathcal{K}\left(\left(T - \lambda I_V\right)^n\right)$$

So, $\mathcal{G}_T(\lambda) \subseteq \mathcal{K}((T - \lambda I_V)^n)$. For the opposite inclusion, suppose $\mathbf{y} \in \mathcal{K}((T - \lambda I_V)^n)$. Then $(T - \lambda I_V)^n(\mathbf{y}) = \mathbf{0}$, so $\mathbf{y} \in \mathcal{G}_T(\lambda)$ and thus $\mathcal{K}((T - \lambda I_V)^n) \subseteq \mathcal{G}_T(\lambda)$. By Definition SE [1604] we have the desired equality of sets.

Theorem GEK [1487] allows us to compute generalized eigenspaces as a single kernel (or null space of a matrix representation) with tools like Theorem KNSI [1311] and Theorem BNS [334]. Also, we do not need to consider all possible powers k and can simply consider the case where k = n. It is worth noting that the "regular" eigenspace is a subspace of the generalized eigenspace since

$$\mathcal{E}_{T}(\lambda) = \mathcal{K}((T - \lambda I_{V})^{1}) \subseteq \mathcal{K}((T - \lambda I_{V})^{n}) = \mathcal{G}_{T}(\lambda)$$

where the subset inclusion is a consequence of Theorem KPLT [1450]. Also, there is no such thing as a "generalized eigenvalue." If λ is not an eigenvalue of T, then the kernel of $T - \lambda I_V$ is trivial and therefore subsequent powers of $T - \lambda I_V$ also have trivial kernels (Theorem KPLT [1450]). So the generalized eigenspace of a scalar that is not already an eigenvalue would be trivial. Alright, we know enough now to compute some generalized eigenspaces. We will record some information about algebraic and geometric multiplicities of eigenvalues (Definition AME [968], Definition GME [969]) as we go, since these observations will be of interest in light of some future theorems.

Example GE4

Generalized eigenspaces, dimension 4 domain

In Example TIS [1476] we presented two invariant subspaces of \mathbb{C}^4 . There was some mystery about just how these were constructed, but we can now reveal that they are generalized

eigenspaces. Example TIS [1476] featured $T: \mathbb{C}^4 \to \mathbb{C}^4$ defined by $T(\mathbf{x}) = A\mathbf{x}$ with A given by

$$A = \begin{bmatrix} -8 & 6 & -15 & 9 \\ -8 & 14 & -10 & 18 \\ 1 & 1 & 3 & 0 \\ 3 & -8 & 2 & -11 \end{bmatrix}$$

A matrix representation of T relative to the standard basis (Definition SUV [409]) will equal A. So we can analyze A with the techniques of Chapter E [944]. Doing so, we find two eigenvalues, $\lambda = 1, -2$, with multiplicities,

$$\alpha_T (1) = 2 \qquad \qquad \gamma_T (1) = 1$$

$$\alpha_T (-2) = 2 \qquad \qquad \gamma_T (-2) = 1$$

To apply Theorem GEK [1487] we subtract each eigenvalue from the diagonal entries of A, raise the result to the power dim (\mathbb{C}^4) = 4, and compute a basis for the null space.

$$\lambda = -2 \qquad (A - (-2)I_4)^4 = \begin{bmatrix} 648 & -1215 & 729 & -1215 \\ -324 & 486 & -486 & 486 \\ -405 & 729 & -486 & 729 \\ 297 & -486 & 405 & -486 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Version 2.99

$$\mathcal{G}_{T}(-2) = \left\langle \left\{ \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\-1\\0\\1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 1 \qquad (A - (1)I_{4})^{4} = \begin{bmatrix} 81 & -405 & -81 & -729\\-108 & -189 & -378 & -486\\-27 & 135 & 27 & 243\\135 & 54 & 351 & 243 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 0 & \frac{7}{3} & 1\\0 & 1 & \frac{2}{3} & 2\\0 & 0 & 0 & 0\\0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{G}_{T}(1) = \left\langle \left\{ \begin{bmatrix} -7\\-2\\3\\0\\0 \end{bmatrix}, \begin{bmatrix} -1\\-2\\0\\1\\1 \end{bmatrix} \right\} \right\rangle$$

In Example TIS [1476] we concluded that these two invariant subspaces formed a direct sum of \mathbb{C}^4 , only at that time, they were called X and W. Now we can write

$$\mathbb{C}^{4}=\mathcal{G}_{T}\left(1\right)\oplus\mathcal{G}_{T}\left(-2\right)$$

This is no accident. Notice that the dimension of each of these invariant subspaces is equal to the algebraic multiplicity of the associated eigenvalue. Not an accident either. (See the upcoming Theorem GESD [1518].) \boxtimes

Example GE6 Generalized eigenspaces, dimension 6 domain Define the linear transformation $S: \mathbb{C}^6 \to \mathbb{C}^6$ by $S(\mathbf{x}) = B\mathbf{x}$ where

$$\begin{bmatrix} 2 & -4 & 25 & -54 & 90 & -37 \\ 2 & -3 & 4 & -16 & 26 & -8 \\ 2 & -3 & 4 & -15 & 24 & -7 \\ 10 & -18 & 6 & -36 & 51 & -2 \\ 8 & -14 & 0 & -21 & 28 & 4 \\ 5 & -7 & -6 & -7 & 8 & 7 \end{bmatrix}$$

Then B will be the matrix representation of S relative to the standard basis (Definition SUV [409]) and we can use the techniques of Chapter E [944] applied to B in order to find the eigenvalues of S.

$$\alpha_{S}(3) = 2 \qquad \qquad \gamma_{S}(3) = 1$$

$$\alpha_{S}(-1) = 4 \qquad \qquad \gamma_{S}(-1) = 2$$

To find the generalized eigenspaces of S we need to subtract an eigenvalue from the diagonal elements of B, raise the result to the power dim (\mathbb{C}^6) = 6 and compute the null space. Here

are the results for the two eigenvalues of S,

If we take the union of the two bases for these two invariant subspaces we obtain the set

$$C = \{\mathbf{v}_1, \, \mathbf{v}_2, \, \mathbf{v}_3, \, \mathbf{v}_4, \, \mathbf{v}_5, \, \mathbf{v}_6\}$$

Version 2.99

You can check that this set is linearly independent (right now we have no guarantee this will happen). Once this is verified, we have a linearly independent set of size 6 inside a vector space of dimension 6, so by Theorem G [852], the set C is a basis for \mathbb{C}^6 . This is enough to apply Theorem DSFB [865] and conclude that

$$\mathbb{C}^{6}=\mathcal{G}_{S}\left(3\right)\oplus\mathcal{G}_{S}\left(-1\right)$$

This is no accident. Notice that the dimension of each of these invariant subspaces is equal to the algebraic multiplicity of the associated eigenvalue. Not an accident either. (See the upcoming Theorem GESD [1518].) \boxtimes

Subsection RLT Restrictions of Linear Transformations

Generalized eigenspaces will prove to be an important type of invariant subspace. A second reason for our interest in invariant subspaces is they provide us with another method for

creating new linear transformations from old ones.

Definition LTR

Linear Transformation Restriction

Suppose that $T: V \to V$ is a linear transformation, and U is an invariant subspace of V relative to T. Define the **restriction** of T to U by

$$T|_{U} \colon U \to U \qquad \qquad T|_{U}\left(\mathbf{u}\right) = T\left(\mathbf{u}\right)$$

(This definition contains Notation LTR.)

It might appear that this definition has not accomplished anything, as $T|_U$ would appear to take on exactly the same values as T. And this is true. However, $T|_U$ differs from Tin the choice of domain and codomain. We tend to give little attention to the domain and codomain of functions, while their defining rules get the spotlight. But the restriction of a linear transformation is all about the choice of domain and codomain. We are *restricting* the rule of the function to a smaller subspace. Notice the importance of only using this construction with an invariant subspace, since otherwise we cannot be assured that the outputs of the function are even contained in the codomain. Maybe this observation should be the key step in the proof of a theorem saying that $T|_U$ is also a linear transformation, but we won't bother.

Example LTRGE

Linear transformation restriction on generalized eigenspace

In order to gain some experience with restrictions of linear transformations, we construct

 \triangle

one and then also construct a matrix representation for the restriction. Furthermore, we will use a generalized eigenspace as the invariant subspace for the construction of the restriction.

Consider the linear transformation $T: \mathbb{C}^5 \to \mathbb{C}^5$ defined by $T(\mathbf{x}) = A\mathbf{x}$, where

$$A = \begin{bmatrix} -22 & -24 & -24 & -24 & -46 \\ 3 & 2 & 6 & 0 & 11 \\ -12 & -16 & -6 & -14 & -17 \\ 6 & 8 & 4 & 10 & 8 \\ 11 & 14 & 8 & 13 & 18 \end{bmatrix}$$

One of the eigenvalues of A is $\lambda = 2$, with geometric multiplicity $\gamma_T(2) = 1$, and algebraic multiplicity $\alpha_T(2) = 3$. We get the generalized eigenspace in the usual manner,

$$W = \mathcal{G}_T (2) = \mathcal{K} \left((T - 2I_{\mathbb{C}^5})^5 \right) = \left\langle \left\{ \begin{bmatrix} -2\\1\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\-1\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} -4\\2\\0\\0\\1 \end{bmatrix} \right\} \right\rangle = \left\langle \left\{ \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3 \right\} \right\rangle$$

By Theorem GESIS [1485], we know W is invariant relative to T, so we can employ Definition LTR [1495] to form the restriction, $T|_W \colon W \to W$.

To better understand exactly what a restriction is (and isn't), we'll form a matrix representation of $T|_W$. This will also be a skill we will use in subsequent examples. For a basis of W we will use $C = {\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3}$. Notice that dim (W) = 3, so our matrix representation will be a square matrix of size 3. Applying Definition MR [1286], we compute

$$\rho_{C}(T(\mathbf{w}_{1})) = \rho_{C}(A\mathbf{w}_{1}) = \rho_{C}\left(\begin{pmatrix} -4\\2\\2\\0\\0\\0 \end{pmatrix}\right) = \rho_{C}\left(2\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix}\right) + 0\begin{bmatrix}0\\-1\\0\\1\\0\\0\end{bmatrix} + 0\begin{bmatrix}-4\\2\\0\\0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}2\\0\\0\\0\\1\end{bmatrix}\right)$$
$$\rho_{C}(T(\mathbf{w}_{2})) = \rho_{C}(A\mathbf{w}_{2}) = \rho_{C}\left(\begin{bmatrix}0\\-2\\2\\2\\2\\-1\end{bmatrix}\right) = \rho_{C}\left(2\begin{bmatrix}-2\\1\\1\\0\\0\\0\end{bmatrix}\right) + 2\begin{bmatrix}0\\-1\\0\\1\\0\\0\end{bmatrix} + (-1)\begin{bmatrix}-4\\2\\0\\0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}2\\2\\-1\end{bmatrix}$$
$$\rho_{C}(T(\mathbf{w}_{3})) = \rho_{C}(A\mathbf{w}_{3}) = \rho_{C}\left(\begin{bmatrix}-6\\3\\-1\\0\\2\end{bmatrix}\right) = \rho_{C}\left((-1)\begin{bmatrix}-2\\1\\1\\0\\0\end{bmatrix} + 0\begin{bmatrix}0\\-1\\0\\1\\0\end{bmatrix}\right) + 2\begin{bmatrix}-4\\2\\0\\0\\1\\0\end{bmatrix}\right) = \begin{bmatrix}-1\\0\\2\end{bmatrix}$$

So the matrix representation of $T|_W$ relative to C is

$$M_{C,C}^{T|_W} = \begin{bmatrix} 2 & 2 & -1 \\ 0 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix}$$

The question arises: how do we use a 3×3 matrix to compute with vectors from \mathbb{C}^5 ? To answer this question, consider the randomly chosen vector

$$\mathbf{w} = \begin{bmatrix} -4\\4\\4\\-2\\-1 \end{bmatrix}$$

First check that $\mathbf{w} \in \mathcal{G}_T(2)$. There are two ways to do this, first verify that

$$(T - 2I_{\mathbb{C}^5})^5 (\mathbf{w}) = (A - 2I_5)^5 \mathbf{w} = \mathbf{0}$$

meeting Definition GES [1484] (with k = 5). Or, express **w** as a linear combination of the basis C for W, to wit, $\mathbf{w} = 4\mathbf{w}_1 - 2\mathbf{w}_2 - \mathbf{w}_3$. Now compute $T|_W(\mathbf{w})$ directly using Definition LTR [1495],

$$T|_{W}(\mathbf{w}) = T(\mathbf{w}) = A\mathbf{w} = \begin{bmatrix} -10\\9\\5\\-4\\0 \end{bmatrix}$$

It was necessary to verify that $\mathbf{w} \in \mathcal{G}_T(2)$, and if we trust our work so far, then this output will also be an element of W, but it would be wise to check this anyway (using either of the methods we used for \mathbf{w}). We'll wait.

Now we will repeat this sample computation, but instead using the matrix representation of $T|_W$ relative to C.

$$T|_{W} (\mathbf{w}) = \rho_{C}^{-1} \left(M_{C,C}^{T|_{W}} \rho_{C} (\mathbf{w}) \right)$$
 Theorem FTMR [1293]
$$= \rho_{C}^{-1} \left(M_{C,C}^{T|_{W}} \rho_{C} (4\mathbf{w}_{1} - 2\mathbf{w}_{2} - \mathbf{w}_{3}) \right)$$

$$= \rho_{C}^{-1} \left(\begin{bmatrix} 2 & 2 & -1 \\ 0 & 2 & 0 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ -1 \end{bmatrix} \right)$$
 Definition VR [1261]
$$= \rho_{C}^{-1} \left(\begin{bmatrix} 5 \\ -4 \\ 0 \end{bmatrix} \right)$$
 Definition MVP [461]
$$= 5\mathbf{w}_{1} - 4\mathbf{w}_{2} + 0\mathbf{w}_{3}$$
 Definition VR [1261]
$$= 5 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + (-4) \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + 0 \begin{bmatrix} -4 \\ 2 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} -10\\9\\5\\-4\\0 \end{bmatrix}$$

which matches the previous computation. Notice how the "action" of $T|_W$ is accomplished by a 3 × 3 matrix multiplying a column vector of size 3. If you would like more practice with these sorts of computations, mimic the above using the other eigenvalue of T, which is $\lambda = -2$. The generalized eigenspace has dimension 2, so the matrix representation of the restriction to the generalized eigenspace will be a 2 × 2 matrix.

Suppose that $T: V \to V$ is a linear transformation and we can find a decomposition of V as a direct sum, say $V = U_1 \oplus U_2 \oplus U_3 \oplus \cdots \oplus U_m$ where each U_i is an invariant subspace of V relative to T. Then, for any $\mathbf{v} \in V$ there is a unique decomposition $\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \cdots + \mathbf{u}_m$ with $\mathbf{u}_i \in U_i, 1 \leq i \leq m$ and furthermore

$$T(\mathbf{v}) = T(\mathbf{u}_{1} + \mathbf{u}_{2} + \mathbf{u}_{3} + \dots + \mathbf{u}_{m})$$

Definition DS [864]
$$= T(\mathbf{u}_{1}) + T(\mathbf{u}_{2}) + T(\mathbf{u}_{3}) + \dots + T(\mathbf{u}_{m})$$

$$= T|_{U_{1}}(\mathbf{u}_{1}) + T|_{U_{2}}(\mathbf{u}_{2}) + T|_{U_{3}}(\mathbf{u}_{3}) + \dots + T|_{U_{m}}(\mathbf{u}_{m})$$

So in a very real sense, we obtain a decomposition of the linear transformation T into the restrictions $T|_{U_i}$, $1 \leq i \leq m$. If we wanted to be more careful, we could extend each restriction to a linear transformation defined on V by setting the output of $T|_{U_i}$ to be the

zero vector for inputs outside of U_i . Then T would be exactly equal to the sum (Definition LTA [1105]) of these extended restrictions. However, the irony of extending our restrictions is more than we could handle right now.

Our real interest is in the matrix representation of a linear transformation when the domain decomposes as a direct sum of invariant subspaces. Consider forming a basis B of V as the union of bases B_i from the individual U_i , i.e. $B = \bigcup_{i=1}^m B_i$. Now form the matrix representation of T relative to B. The result will be block diagonal, where each block is the matrix representation of a restriction $T|_{U_i}$ relative to a basis B_i , $M_{B_i,B_i}^{T|_{U_i}}$. Though we did not have the definitions to describe it then, this is exactly what was going on in the latter portion of the proof of Theorem CFNLT [1458]. Two examples should help to clarify these ideas.

Example ISMR4

Invariant subspaces, matrix representation, dimension 4 domain

Example TIS [1476] and Example GE4 [1488] describe a basis of \mathbb{C}^4 which is derived from bases for two invariant subspaces (both generalized eigenspaces). In this example we will construct a matrix representation of the linear transformation T relative to this basis. Recycling the notation from Example TIS [1476], we work with the basis,

$$B = \{\mathbf{w}_1, \, \mathbf{w}_2, \, \mathbf{x}_1, \, \mathbf{x}_2\} = \left\{ \begin{bmatrix} -7\\ -2\\ 3\\ 0 \end{bmatrix}, \, \begin{bmatrix} -1\\ -2\\ 0\\ 1 \end{bmatrix}, \, \begin{bmatrix} -3\\ -1\\ 1\\ 0\\ 1 \end{bmatrix}, \, \begin{bmatrix} 0\\ -1\\ 0\\ 1 \end{bmatrix} \right\}$$

Now we compute the matrix representation of T relative to B, borrowing some computations from Example TIS [1476],

$$\rho_B \left(T \left(\mathbf{w}_1 \right) \right) = \rho_B \left(\begin{bmatrix} -1 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right) = \rho_B \left((0) \mathbf{w}_1 + (1) \mathbf{w}_2 \right) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{w}_2 \right) \right) = \rho_B \left(\begin{bmatrix} 5 \\ -2 \\ -3 \\ 2 \end{bmatrix} \right) = \rho_B \left((-1) \mathbf{w}_1 + (2) \mathbf{w}_2 \right) = \begin{bmatrix} -1 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{x}_1 \right) \right) = \rho_B \left(\begin{bmatrix} 3 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right) = \rho_B \left((-1) \mathbf{x}_1 + (1) \mathbf{x}_2 \right) = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$
$$\rho_B \left(T \left(\mathbf{x}_2 \right) \right) = \rho_B \left(\begin{bmatrix} 3 \\ 4 \\ -1 \\ -3 \end{bmatrix} \right) = \rho_B \left((-1) \mathbf{x}_1 + (-3) \mathbf{x}_2 \right) = \begin{bmatrix} 0 \\ 0 \\ -1 \\ -1 \\ -3 \end{bmatrix}$$

Applying Definition MR [1286], we have

$$M_{B,B}^{T} = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & -3 \end{bmatrix}$$

The interesting feature of this representation is the two 2×2 blocks on the diagonal that arise from the decomposition of \mathbb{C}^4 into a direct sum (of generalized eigenspaces). Or maybe the interesting feature of this matrix is the two 2×2 submatrices in the "other" corners that are all zero. You decide.

Example ISMR6 Invariant subspaces, matrix representation, dimension 6 domain In Example GE6 [1490] we computed the generalized eigenspaces of the linear

In Example GE6 [1490] we computed the generalized eigenspaces of the linear transformation $S: \mathbb{C}^6 \to \mathbb{C}^6$ by $S(\mathbf{x}) = B\mathbf{x}$ where

$$\begin{bmatrix} 2 & -4 & 25 & -54 & 90 & -37 \\ 2 & -3 & 4 & -16 & 26 & -8 \\ 2 & -3 & 4 & -15 & 24 & -7 \\ 10 & -18 & 6 & -36 & 51 & -2 \\ 8 & -14 & 0 & -21 & 28 & 4 \\ 5 & -7 & -6 & -7 & 8 & 7 \end{bmatrix}$$

From this we found the basis

of \mathbb{C}^6 where $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of $\mathcal{G}_S(3)$ and $\{\mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is a basis of $\mathcal{G}_S(-1)$. We can employ C in the construction of a matrix representation of S (Definition MR [1286]). Here are the computations,

$$\rho_C\left(S\left(\mathbf{v}_1\right)\right) = \rho_C\left(\begin{bmatrix}11\\3\\3\\7\\4\\1\end{bmatrix}\right) = \rho_C\left(4\mathbf{v}_1 + 1\mathbf{v}_2\right) = \begin{bmatrix}4\\1\\0\\0\\0\\0\end{bmatrix}$$

$$\rho_{C} \left(S \left(\mathbf{v}_{2} \right) \right) = \rho_{C} \left(\begin{bmatrix} -14\\ -3\\ -4\\ -1\\ 2 \end{bmatrix} \right) = \rho_{C} \left((-1)\mathbf{v}_{1} + 2\mathbf{v}_{2} \right) = \begin{bmatrix} -1\\ 2\\ 0\\ 0\\ 0\\ 0 \end{bmatrix}$$

$$\rho_{C} \left(S \left(\mathbf{v}_{3} \right) \right) = \rho_{C} \left(\begin{bmatrix} 23\\ 5\\ 2\\ -2\\ -2\\ -2 \end{bmatrix} \right) = \rho_{C} \left(5\mathbf{v}_{3} + 2\mathbf{v}_{4} + (-2)\mathbf{v}_{5} + (-2)\mathbf{v}_{6} \right) = \begin{bmatrix} 0\\ 0\\ 5\\ 2\\ -2\\ -2\\ -2 \end{bmatrix}$$

$$\rho_{C} \left(S \left(\mathbf{v}_{4} \right) \right) = \rho_{C} \left(\begin{bmatrix} -46\\ -11\\ -10\\ -2\\ 5\\ 4 \end{bmatrix} \right) = \rho_{C} \left((-10)\mathbf{v}_{3} + (-2)\mathbf{v}_{4} + 5\mathbf{v}_{5} + 4\mathbf{v}_{6} \right) = \begin{bmatrix} 0\\ 0\\ -10\\ -2\\ 5\\ 4 \end{bmatrix}$$

$$\rho_{C}\left(S\left(\mathbf{v}_{5}\right)\right) = \rho_{C}\left(\begin{bmatrix}78\\19\\1\\-1\\-10\\-7\end{bmatrix}\right) = \rho_{C}\left(17\mathbf{v}_{3} + 1\mathbf{v}_{4} + (-10)\mathbf{v}_{5} + (-7)\mathbf{v}_{6}\right) = \begin{bmatrix}0\\0\\17\\1\\-10\\-7\end{bmatrix}$$

$$\rho_{C}\left(S\left(\mathbf{v}_{6}\right)\right) = \rho_{C}\left(\begin{bmatrix}-35\\-9\\-8\\2\\6\\3\end{bmatrix}\right) = \rho_{C}\left((-8)\mathbf{v}_{3} + 2\mathbf{v}_{4} + 6\mathbf{v}_{5} + 3\mathbf{v}_{6}\right) = \begin{bmatrix}0\\0\\-8\\2\\6\\3\end{bmatrix}$$

These column vectors are the columns of the matrix representation, so we obtain

$$M_{C,C}^{S} = \begin{bmatrix} 4 & -1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & -10 & 17 & -8 \\ 0 & 0 & 2 & -2 & 1 & 2 \\ 0 & 0 & -2 & 5 & -10 & 6 \\ 0 & 0 & -2 & 4 & -7 & 3 \end{bmatrix}$$

As before, the key feature of this representation is the 2×2 and 4×4 blocks on the diagonal. We will discover in the final theorem of this section (Theorem RGEN [1507]) that

we already understand these blocks fairly well. For now, we recognize them as arising from generalized eigenspaces and suspect that their sizes are equal to the algebraic multiplicities of the eigenvalues. \boxtimes

The paragraph prior to these last two examples is worth repeating. A basis derived from a direct sum decomposition into invariant subspaces will provide a matrix representation of a linear transformation with a block diagonal form.

Diagonalizing a linear transformation is the most extreme example of decomposing a vector space into invariant subspaces. When a linear transformation is diagonalizable, then there is a basis composed of eigenvectors (Theorem DC [1039]). Each of these basis vectors can be used individually as the lone element of a spanning set for an invariant subspace (Theorem EIS [1479]). So the domain decomposes into a direct sum of one-dimensional invariant subspaces (Theorem DSFB [865]). The corresponding matrix representation is then block diagonal with all the blocks of size 1, i.e. the matrix is diagonal. Section NLT [1435], Section IS [1475] and Section JCF [1517] are all devoted to generalizing this extreme situation when there are not enough eigenvectors available to make such a complete decomposition and arrive at such an elegant matrix representation.

One last theorem will roll up much of this section and Section NLT [1435] into one nice, neat package.

Theorem RGEN Restriction to Generalized Eigenspace is Nilpotent

Suppose $T: V \to V$ is a linear transformation with eigenvalue λ . Then the linear transfor-

mation $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$ is nilpotent.

Proof Notice first that every subspace of V is invariant with respect to I_V , so $I_{\mathcal{G}_T(\lambda)} = I_V|_{\mathcal{G}_T(\lambda)}$. Let $n = \dim(V)$ and choose $\mathbf{v} \in \mathcal{G}_T(\lambda)$. Then

$$(T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)})^n (\mathbf{v}) = (T - \lambda I_V)^n (\mathbf{v})$$
 Definition LTR [1495]
= 0 Theorem GEK [1487]

So by Definition NLT [1437], $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$ is nilpotent.

The proof of Theorem RGEN [1507] indicates that the index of the nilpotent linear transformation is less than or equal to the dimension of V. In practice, it will be less than or equal to the dimension of the linear transformation, $\mathcal{G}_T(\lambda)$. In any event, the exact value of this index will be of some interest, so we define it now. Notice that this is a property of the eigenvalue λ , similar to the algebraic and geometric multiplicities (Definition AME [968], Definition GME [969]).

Definition IE

Index of an Eigenvalue

Suppose $T: V \to V$ is a linear transformation with eigenvalue λ . Then the **index** of λ , $\iota_T(\lambda)$, is the index of the nilpotent linear transformation $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$. (This definition contains Notation IE.)

Example GENR6

Generalized eigenspaces and nilpotent restrictions, dimension 6 domain

In Example GE6 [1490] we computed the generalized eigenspaces of the linear transformation

 \square

 $S: \mathbb{C}^6 \to \mathbb{C}^6$ defined by $S(\mathbf{x}) = B\mathbf{x}$ where

$$\begin{bmatrix} 2 & -4 & 25 & -54 & 90 & -37 \\ 2 & -3 & 4 & -16 & 26 & -8 \\ 2 & -3 & 4 & -15 & 24 & -7 \\ 10 & -18 & 6 & -36 & 51 & -2 \\ 8 & -14 & 0 & -21 & 28 & 4 \\ 5 & -7 & -6 & -7 & 8 & 7 \end{bmatrix}$$

The generalized eigenspace, $\mathcal{G}_S(3)$, has dimension 2, while $\mathcal{G}_S(-1)$, has dimension 4. We'll investigate each thoroughly in turn, with the intent being to illustrate Theorem RGEN [1507]. Much of our computations will be repeats of those done in Example ISMR6 [1503].

For $U = \mathcal{G}_S(3)$ we compute a matrix representation of $S|_U$ using the basis found in Example GE6 [1490],

$$B = \{\mathbf{u}_1, \, \mathbf{u}_2\} = \left\{ \begin{bmatrix} 4\\1\\1\\1\\2\\1\\0\end{bmatrix}, \begin{bmatrix} -5\\-1\\-1\\-1\\-1\\0\\1\end{bmatrix} \right\}$$

Since B has size 2, we obtain a 2×2 matrix representation (Definition MR [1286]) from

$$\rho_B\left(S|_U\left(\mathbf{u}_1\right)\right) = \rho_B\left(\begin{bmatrix}11\\3\\7\\4\\1\end{bmatrix}\right) = \rho_B\left(4\mathbf{u}_1 + \mathbf{u}_2\right) = \begin{bmatrix}4\\1\end{bmatrix}$$
$$\rho_B\left(S|_U\left(\mathbf{u}_2\right)\right) = \rho_B\left(\begin{bmatrix}-14\\-3\\-3\\-4\\-1\\2\end{bmatrix}\right) = \rho_B\left((-1)\mathbf{u}_1 + 2\mathbf{u}_2\right) = \begin{bmatrix}-1\\2\end{bmatrix}$$

Thus

$$M = M_{U,U}^{S|_U} = \begin{bmatrix} 4 & -1\\ 1 & 2 \end{bmatrix}$$

Now we can illustrate Theorem RGEN [1507] with powers of the matrix representation (rather than the restriction itself),

$$M - 3I_2 = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \qquad (M - 3I_2)^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

So $M - 3I_2$ is a nilpotent matrix of index 2 (meaning that $S|_U - 3I_U$ is a nilpotent linear transformation of index 2) and according to Definition IE [1508] we say $\iota_S(3) = 2$.

For $W = \mathcal{G}_S(-1)$ we compute a matrix representation of $S|_W$ using the basis found in Example GE6 [1490],

$$C = \{\mathbf{w}_1, \, \mathbf{w}_2, \, \mathbf{w}_3, \, \mathbf{w}_4\} = \left\{ \begin{bmatrix} 5\\3\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix} -2\\-3\\0\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix} 4\\5\\0\\0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} -5\\-3\\0\\0\\0\\1\\0\end{bmatrix} \right\}$$

Since C has size 4, we obtain a 4×4 matrix representation (Definition MR [1286]) from

$$\rho_C \left(S|_W \left(\mathbf{w}_1 \right) \right) = \rho_C \left(\begin{bmatrix} 23\\5\\2\\-2\\-2\\-2 \end{bmatrix} \right) = \rho_C \left(5\mathbf{w}_1 + 2\mathbf{w}_2 + (-2)\mathbf{w}_3 + (-2)\mathbf{w}_4 \right) = \begin{bmatrix} 5\\2\\-2\\-2\\-2 \end{bmatrix}$$

$$\rho_{C}\left(S|_{W}\left(\mathbf{w}_{2}\right)\right) = \rho_{C}\left(\begin{bmatrix}-46\\-11\\-10\\-2\\5\\4\end{bmatrix}\right) = \rho_{C}\left((-10)\mathbf{w}_{1} + (-2)\mathbf{w}_{2} + 5\mathbf{w}_{3} + 4\mathbf{w}_{4}\right) = \begin{bmatrix}-10\\-2\\5\\4\end{bmatrix}$$
$$\rho_{C}\left(S|_{W}\left(\mathbf{w}_{3}\right)\right) = \rho_{C}\left(\begin{bmatrix}78\\19\\17\\1\\-10\\-7\end{bmatrix}\right) = \rho_{C}\left(17\mathbf{w}_{1} + \mathbf{w}_{2} + (-10)\mathbf{w}_{3} + (-7)\mathbf{w}_{4}\right) = \begin{bmatrix}17\\1\\-10\\-7\end{bmatrix}$$
$$\rho_{C}\left(S|_{W}\left(\mathbf{w}_{4}\right)\right) = \rho_{C}\left(\begin{bmatrix}-35\\-9\\-8\\2\\6\\3\end{bmatrix}\right) = \rho_{C}\left((-8)\mathbf{w}_{1} + 2\mathbf{w}_{2} + 6\mathbf{w}_{3} + 3\mathbf{w}_{4}\right) = \begin{bmatrix}-8\\2\\6\\3\end{bmatrix}$$

Thus

$$N = M_{W,W}^{S|_W} = \begin{bmatrix} 5 & -10 & 17 & -8\\ 2 & -2 & 1 & 2\\ -2 & 5 & -10 & 6\\ -2 & 4 & -7 & 3 \end{bmatrix}$$

Now we can illustrate Theorem RGEN [1507] with powers of the matrix representation (rather than the restriction itself),

So $N - (-1)I_4$ is a nilpotent matrix of index 3 (meaning that $S|_W - (-1)I_W$ is a nilpotent linear transformation of index 3) and according to Definition IE [1508] we say $\iota_S(-1) = 3$.

Notice that if we were to take the union of the two bases of the generalized eigenspaces, we would have a basis for \mathbb{C}^6 . Then a matrix representation of S relative to this basis would be the same block diagonal matrix we found in Example ISMR6 [1503], only we now understand each of these blocks as being very close to being a nilpotent matrix. \boxtimes

Invariant subspaces, and restrictions of linear transformations, are topics you will see again and again if you continue with further study of linear algebra. Our reasons for discussing them now is to arrive at a nice matrix representation of the restriction of a linear transformation to one of its generalized eigenspaces. Here's the theorem.

Theorem MRRGE

Matrix Representation of a Restriction to a Generalized Eigenspace

Suppose that $T: V \to V$ is a linear transformation with eigenvalue λ . Then there is a basis of the generalized eigenspace $\mathcal{G}_T(\lambda)$ such that the restriction $T|_{\mathcal{G}_T(\lambda)}: \mathcal{G}_T(\lambda) \to \mathcal{G}_T(\lambda)$ has a matrix representation that is block diagonal where each block is a Jordan block of the form $J_n(\lambda)$.

Proof Theorem RGEN [1507] tells us that $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$ is a nilpotent linear transformation. Theorem CFNLT [1458] tells us that a nilpotent linear transformation has a basis for its domain that yields a matrix representation that is block diagonal where the blocks are Jordan blocks of the form $J_n(0)$. Let *B* be a basis of $\mathcal{G}_T(\lambda)$ that yields such a matrix representation for $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$.

By Definition LTA [1105], we can write

$$T|_{\mathcal{G}_T(\lambda)} = \left(T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}\right) + \lambda I_{\mathcal{G}_T(\lambda)}$$

The matrix representation of $\lambda I_{\mathcal{G}_T(\lambda)}$ relative to the basis *B* is then simply the diagonal matrix λI_m , where $m = \dim(\mathcal{G}_T(\lambda))$. By Theorem MRSLT [1300] we have the rather unwieldy expression,

$$M_{B,B}^{T|_{\mathcal{G}_{T}(\lambda)}} = M_{B,B}^{\left(T|_{\mathcal{G}_{T}(\lambda)} - \lambda I_{\mathcal{G}_{T}(\lambda)}\right) + \lambda I_{\mathcal{G}_{T}(\lambda)}}$$
$$= M_{B,B}^{T|_{\mathcal{G}_{T}(\lambda)} - \lambda I_{\mathcal{G}_{T}(\lambda)}} + M_{B,B}^{I_{\mathcal{G}_{T}(\lambda)}}$$

The first of these matrix representations has Jordan blocks with zero in every diagonal entry, while the second matrix representation has λ in every diagonal entry. The result of adding the two representations is to convert the Jordan blocks from the form $J_n(0)$ to the form $J_n(\lambda)$.

Of course, Theorem CFNLT [1458] provides some extra information on the sizes of the Jordan blocks in a representation and we could carry over this information to Theorem MRRGE [1514], but will save that for a subsequent application of this result.

Subsection EXC Exercises

T10 Suppose that $T: V \to V$ is linear transformation, and p(x) is a polynomial. Then define the new linear transformation $p(T): V \to V$ by interpreting the coefficients of the terms of the polynomial as scalar multiples of linear transformations (Definition LTSM [1108]), addition of terms as the sum of linear transformations (Definition LTA [1105]), and powers as repeated composition of linear transformations (Definition LTC [1111]). Prove that $T \circ p(T) = p(T) \circ T$.

Use this observation to give a shorter argument for the proof of the invariance of the generalized eigenspace in Theorem GESIS [1485]. Contributed by Robert Beezer

Section JCF Jordan Canonical Form

This section is in draft form Needs examples near beginning

We have seen in Section IS [1475] that generalized eigenspaces are invariant subspaces that in every instance have led to a direct sum decomposition of the domain of the associated linear transformation. This allows us to create a block diagonal matrix representation (Example ISMR4 [1501], Example ISMR6 [1503]). We also know from Theorem RGEN [1507] that the restriction of a linear transformation to a generalized eigenspace is almost a nilpotent linear transformation. Of course, we understand nilpotent linear transformations very well from Section NLT [1435] and we have carefully determined a nice matrix representation for them.

So here is the game plan for the final push. Prove that the domain of a linear transformation always decomposes into a direct sum of generalized eigenspaces. We have unravelled Theorem RGEN [1507] at Theorem MRRGE [1514] so that we can formulate the matrix representations of the restrictions on the generalized eigenspaces using our storehouse of results about nilpotent linear transformations. Arrive at a matrix representation of *any* linear transformation that is block diagonal with each block being a Jordan block.

Subsection GESD Generalized Eigenspace Decomposition

In Theorem UTMR [1419] we were able to show that any linear transformation from V to V has an upper triangular matrix representation (Definition UTM [1415]). We will now show that we can improve on the basis yielding this representation by massaging the basis so that the matrix representation is also block diagonal. The subspaces associated with each block will be generalized eigenspaces, so the most general result will be a decomposition of the domain of a linear transformation into a direct sum of generalized eigenspaces.

Theorem GESD

Generalized Eigenspace Decomposition

Suppose that $T: V \to V$ is a linear transformation with distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_m$. Then

$$V = \mathcal{G}_{T}(\lambda_{1}) \oplus \mathcal{G}_{T}(\lambda_{2}) \oplus \mathcal{G}_{T}(\lambda_{3}) \oplus \cdots \oplus \mathcal{G}_{T}(\lambda_{m})$$

Proof Suppose that dim (V) = n and the *n* (not necessarily distinct) eigenvalues of *T* are $\rho_1, \rho_2, \rho_3, \ldots, \rho_n$. We begin with a basis of *V* that yields an upper triangular matrix representation, as guaranteed by Theorem UTMR [1419], $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$. Since the matrix representation is upper triangular, and the eigenvalues of the linear transformation

 \square

are the diagonal elements we can choose this basis so that there are then scalars a_{ij} , $1 \le j \le n$, $1 \le i \le j - 1$ such that

$$\Gamma(\mathbf{x}_j) = \sum_{i=1}^{j-1} a_{ij} \mathbf{x}_i + \rho_j \mathbf{x}_j$$

We now define a new basis for V which is just a slight variation in the basis B. Choose any k and ℓ such that $1 \leq k < \ell \leq n$ and $\rho_k \neq \rho_\ell$. Define the scalar $\alpha = a_{kl}/(\rho_\ell - \rho_k)$. The new basis is $C = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_n\}$ where

$$\mathbf{y}_j = \mathbf{x}_j, \quad j \neq \ell, \ 1 \le j \le n$$
 $\mathbf{y}_\ell = \mathbf{x}_\ell + \alpha \mathbf{x}_k$

We now compute the values of the linear transformation T with inputs from C, noting carefully the changed scalars in the linear combinations of C describing the outputs. These changes will translate to minor changes in the matrix representation built using the basis C. There are three cases to consider, depending on which column of the matrix representation we are examining. First, assume $j < \ell$. Then

$$T (\mathbf{y}_j) = T (\mathbf{x}_j)$$

= $\sum_{i=1}^{j-1} a_{ij} \mathbf{x}_i + \rho_j \mathbf{x}_j$
= $\sum_{i=1}^{j-1} a_{ij} \mathbf{y}_i + \rho_j \mathbf{y}_j$

That seems a bit pointless. The first $\ell - 1$ columns of the matrix representations of T relative to B and C are identical. OK, if that was too easy, here's the main act. Assume $j = \ell$. Then

$$T (\mathbf{y}_{\ell}) = T (\mathbf{x}_{\ell} + \alpha \mathbf{x}_{k})$$

$$= T (\mathbf{x}_{\ell}) + \alpha T (\mathbf{x}_{k})$$

$$= \left(\sum_{i=1}^{\ell-1} a_{i\ell} \mathbf{x}_{i} + \rho_{\ell} \mathbf{x}_{\ell}\right) + \alpha \left(\sum_{i=1}^{k-1} a_{ik} \mathbf{x}_{i} + \rho_{k} \mathbf{x}_{k}\right)$$

$$= \sum_{i=1}^{\ell-1} a_{i\ell} \mathbf{x}_{i} + \rho_{\ell} \mathbf{x}_{\ell} + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{x}_{i} + \alpha \rho_{k} \mathbf{x}_{k}$$

$$= \sum_{i=1}^{\ell-1} a_{i\ell} \mathbf{x}_{i} + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{x}_{i} + \alpha \rho_{k} \mathbf{x}_{k} + \rho_{\ell} \mathbf{x}_{\ell}$$

$$= \sum_{\substack{i=1\\i \neq k}}^{\ell-1} a_{i\ell} \mathbf{x}_{i} + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{x}_{i} + a_{kl} \mathbf{x}_{k} + \alpha \rho_{k} \mathbf{x}_{k} + \rho_{\ell} \mathbf{x}_{\ell}$$

$$= \sum_{\substack{i=1\\i \neq k}}^{\ell-1} a_{i\ell} \mathbf{x}_{i} + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{x}_{i} + a_{kl} \mathbf{x}_{k} + \alpha \rho_{k} \mathbf{x}_{k} - \rho_{\ell} \alpha \mathbf{x}_{k} + \rho_{\ell} \mathbf{x}_{\ell}$$

$$= \sum_{\substack{i=1\\i\neq k}}^{\ell-1} a_{i\ell} \mathbf{x}_i + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{x}_i + (a_{kl} + \alpha \rho_k - \rho_\ell \alpha) \mathbf{x}_k + \rho_\ell (\alpha \mathbf{x}_k + \mathbf{x}_\ell)$$
$$= \sum_{\substack{i=1\\i\neq k}}^{\ell-1} a_{i\ell} \mathbf{x}_i + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{x}_i + (a_{kl} + \alpha (\rho_k - \rho_\ell)) \mathbf{x}_k + \rho_\ell (\mathbf{x}_\ell + \alpha \mathbf{x}_k)$$
$$= \sum_{\substack{i=1\\i\neq k}}^{\ell-1} a_{i\ell} \mathbf{y}_i + \sum_{i=1}^{k-1} \alpha a_{ik} \mathbf{y}_i + (a_{kl} + \alpha (\rho_k - \rho_\ell)) \mathbf{y}_k + \rho_\ell \mathbf{y}_\ell$$

So how different are the matrix representations relative to B and C in column ℓ ? For i > k, the coefficient of \mathbf{y}_i is a_{ij} , as in the representation relative to B. It is a different story for $i \leq k$, where the coefficients of \mathbf{y}_i may be very different. We are especially interested in the coefficient of \mathbf{y}_k . In fact, this whole first part of this proof is about this particular entry of the matrix representation. The coefficient of \mathbf{y}_k is

$$a_{kl} + \alpha \left(\rho_k - \rho_\ell\right) = a_{kl} + \frac{a_{kl}}{\rho_\ell - \rho_k} \left(\rho_k - \rho_\ell\right)$$
$$= a_{kl} + (-1)a_{kl}$$
$$= 0$$

If the definition of α was a mystery, then no more. In the matrix representation of T relative to C, the entry in column ℓ , row k is a zero. Nice. The only price we pay is that other entries in column ℓ , specifically rows 1 through k - 1, may also change in a way we can't control.

One more case to consider. Assume $j > \ell$. Then

T

$$\begin{aligned} (\mathbf{y}_j) &= T\left(\mathbf{x}_j\right) \\ &= \sum_{\substack{i=1\\i\neq\ell,k}}^{j-1} a_{ij}\mathbf{x}_i + \rho_j \mathbf{x}_j \\ &= \sum_{\substack{i=1\\i\neq\ell,k}}^{j-1} a_{ij}\mathbf{x}_i + a_{\ell j}\mathbf{x}_\ell + a_{kj}\mathbf{x}_k + \rho_j \mathbf{x}_j \\ &= \sum_{\substack{i=1\\i\neq\ell,k}}^{j-1} a_{ij}\mathbf{x}_i + a_{\ell j}(\mathbf{x}_\ell + \alpha \mathbf{x}_k) + (a_{kj} - \alpha a_{\ell j})\mathbf{x}_k + \rho_j \mathbf{x}_j \\ &= \sum_{\substack{i=1\\i\neq\ell,k}}^{j-1} a_{ij}\mathbf{y}_i + a_{\ell j}(\mathbf{x}_\ell + \alpha \mathbf{x}_k) + (a_{kj} - \alpha a_{\ell j})\mathbf{x}_k + \rho_j \mathbf{x}_j \end{aligned}$$

As before, we ask: how different are the matrix representations relative to B and C in column j? Only \mathbf{y}_k has a coefficient different from the corresponding coefficient when the basis is B. So in the matrix representations, the only entries to change are in row k, for columns $\ell + 1$ through n.

What have we accomplished? With a change of basis, we can place a zero in a desired entry (row k, column ℓ) of the matrix representation, leaving most of the entries untouched. The only entries to possibly change are above the new zero entry, or to the right of the new zero entry. Suppose we repeat this procedure, starting by "zeroing out" the entry above the diagonal in the second column and first row. Then we move right to the third column, and zero out the element just above the diagonal in the second row. Next we zero out the element in the third column and first row. Then tackle the fourth column, work upwards from the diagonal, zeroing out elements as we go. Entries above, and to the right will repeatedly change, but newly created zeros will never get wrecked, since they are below, or just to the left of the entry we are working on. Similarly the values on the diagonal do not change either. This entire argument can be retooled in the language of change-of-basis matrices and similarity transformations, and this is the approach taken by Noble in his Applied Linear Algebra. It is interesting to concoct the change-of-basis matrix between the matrices B and C and compute the inverse.

Perhaps you have noticed that we have to be just a bit more careful than the previous paragraph suggests. The definition of α has a denominator that cannot be zero, which restricts our maneuvers to zeroing out entries in row k and column ℓ only when $\rho_k \neq \rho_\ell$. So we do not necessarily arrive at a diagonal matrix. More carefully we can write

$$T(\mathbf{y}_j) = \sum_{\substack{i=1\\i:\,\rho_i=\rho_j}}^{j-1} b_{ij}\mathbf{y}_i + \rho_j\mathbf{y}_j$$

where the b_{ij} are our new coefficients after repeated changes, the \mathbf{y}_i are the new basis

vectors, and the condition " $i : \rho_i = \rho_j$ " means that we only have terms in the sum involving vectors whose final coefficients are identical diagonal values (the eigenvalues). Now reorder the basis vectors carefully. Group together vectors that have equal diagonal entries in the matrix representation, but within each group preserve the order of the precursor basis. This grouping will create a block diagonal structure for the matrix representation, while otherwise preserving the order of the basis will retain the upper triangular form of the representation. So we can arrive at a basis that yields a matrix representation that is upper triangular and block diagonal, with the diagonal entries of each block all equal to a common eigenvalue of the linear transformation.

More carefully, employing the distinct eigenvalues of T, λ_i , $1 \leq i \leq m$, we can assert there is a set of basis vectors for V, \mathbf{u}_{ij} , $1 \leq i \leq m$, $1 \leq j \leq \alpha_T(\lambda_i)$, such that

$$T\left(\mathbf{u}_{ij}\right) = \sum_{k=1}^{j-1} b_{ijk} \mathbf{u}_{ik} + \lambda_i \mathbf{u}_{ij}$$

So the subspace $U_i = \langle \{ \mathbf{u}_{ij} | 1 \leq j \leq \alpha_T(\lambda_i) \} \rangle$, $1 \leq i \leq m$ is an invariant subspace of V relative to T and the restriction $T|_{U_i}$ has an upper triangular matrix representation relative to the basis $\{ \mathbf{u}_{ij} | 1 \leq j \leq \alpha_T(\lambda_i) \}$ where the diagonal entries are all equal to λ_i . Notice too that with this definition,

$$V = U_1 \oplus U_2 \oplus U_3 \oplus \cdots \oplus U_m$$

Whew. This is a good place to take a break, grab a cup of coffee, use the toilet, or go for a short stroll, before we show that U_i is a subspace of the generalized eigenspace $\mathcal{G}_T(\lambda_i)$. This

will follow if we can prove that each of the basis vectors for U_i is a generalized eigenvector of T for λ_i (Definition GEV [1484]). We need some power of $T - \lambda_i I_V$ that takes \mathbf{u}_{ij} to the zero vector. We prove by induction on j (Technique I [1630]) the claim that $(T - \lambda_i I_V)^j$ (\mathbf{u}_{ij}) = **0**. For j = 1 we have,

$$(T - \lambda_i I_V) (\mathbf{u}_{i1}) = T (\mathbf{u}_{i1}) - \lambda_i I_V (\mathbf{u}_{i1})$$
$$= T (\mathbf{u}_{i1}) - \lambda_i \mathbf{u}_{i1}$$
$$= \lambda_i \mathbf{u}_{i1} - \lambda_i \mathbf{u}_{i1}$$
$$= \mathbf{0}$$

For the induction step, assume that if k < j, then $(T - \lambda_i I_V)^k$ takes \mathbf{u}_{ik} to the zero vector. Then

$$(T - \lambda_i I_V)^j (\mathbf{u}_{ij}) = (T - \lambda_i I_V)^{j-1} ((T - \lambda_i I_V) (\mathbf{u}_{ij}))$$

$$= (T - \lambda_i I_V)^{j-1} (T (\mathbf{u}_{ij}) - \lambda_i I_V (\mathbf{u}_{ij}))$$

$$= (T - \lambda_i I_V)^{j-1} (T (\mathbf{u}_{ij}) - \lambda_i \mathbf{u}_{ij})$$

$$= (T - \lambda_i I_V)^{j-1} \left(\sum_{k=1}^{j-1} b_{ijk} \mathbf{u}_{ik} + \lambda_i \mathbf{u}_{ij} - \lambda_i \mathbf{u}_{ij}\right)$$

$$= (T - \lambda_i I_V)^{j-1} \left(\sum_{k=1}^{j-1} b_{ijk} \mathbf{u}_{ik}\right)$$

$$= \sum_{k=1}^{j-1} b_{ijk} (T - \lambda_i I_V)^{j-1} (\mathbf{u}_{ik})$$

$$= \sum_{k=1}^{j-1} b_{ijk} (T - \lambda_i I_V)^{j-1-k} ((T - \lambda_i I_V)^k (\mathbf{u}_{ik}))$$

$$= \sum_{k=1}^{j-1} b_{ijk} (T - \lambda_i I_V)^{j-1-k} (\mathbf{0})$$

$$= \sum_{k=1}^{j-1} b_{ijk} \mathbf{0}$$

$$= \mathbf{0}$$

This completes the induction step. Since every vector of the spanning set for U_i is an element of the subspace $\mathcal{G}_T(\lambda_i)$, Property AC [660] and Property SC [660] allow us to conclude that $U_i \subseteq \mathcal{G}_T(\lambda_i)$. Then by Definition S [694], U_i is a subspace of $\mathcal{G}_T(\lambda_i)$. Notice that this inductive proof could be interpreted to say that every element of U_i is a generalized eigenvector of T for λ_i , and the algebraic multiplicity of λ_i is a sufficiently high power to demonstrate this via the definition for each vector.

We are now prepared for our final argument in this long proof. We wish to establish that the dimension of the subspace $\mathcal{G}_T(\lambda_i)$ is the algebraic multiplicity of λ_i . This will be enough to show that U_i and $\mathcal{G}_T(\lambda_i)$ are equal, and will finally provide the desired direct sum decomposition.

We will prove by induction (Technique I [1630]) the following claim. Suppose that $T: V \to V$ is a linear transformation and B is a basis for V that provides an upper triangular matrix representation of T. The number of times any eigenvalue λ occurs on the diagonal of the representation is greater than or equal to the dimension of the generalized eigenspace $\mathcal{G}_T(\lambda)$.

We will use the symbol m for the dimension of V so as to avoid confusion with our notation for the nullity. So dim V = m and our proof will proceed by induction on m. Use the notation $\#_T(\lambda)$ to count the number of times λ occurs on the diagonal of a matrix representation of T. We want to show that

$$\#_{T}(\lambda) \geq \dim \left(\mathcal{G}_{T}(\lambda)\right)$$

= dim $\left(\mathcal{K}((T-\lambda)^{m})\right)$ Theorem GEK [1487]
= $n\left((T-\lambda)^{m}\right)$ Definition NOLT [1234]

For the base case, dim V = 1. Every matrix representation of T is an upper triangular matrix with the lone eigenvalue of T, λ , as the single diagonal entry. So $\#_T(\lambda) = 1$. The generalized eigenspace of λ is not trivial (since by Theorem GEK [1487] it equals the regular eigenspace), so it cannot be a subspace of dimension zero, and thus dim $(\mathcal{G}_T(\lambda)) = 1$.

Now for the induction step, assume the claim is true for any linear transformation defined on a vector space with dimension m-1 or less. Suppose that $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m}$ is a basis for V that yields an upper triangular matrix representation for T with diagonal entries $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_m$. Then $U = \langle {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{m-1}} \rangle$ is a subspace of V that is invariant relative to T. The restriction $T|_U: U \to U$ is then a linear transformation defined on U, a vector space of dimension m-1. A matrix representation of $T|_U$ relative to the basis $C = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{m-1}\}$ will be an upper triangular matrix with diagonal entries $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{m-1}$. We can therefore apply the induction hypothesis to $T|_U$ and its representation relative to C.

Suppose that λ is any eigenvalue of T. Then suppose that $\mathbf{v} \in \mathcal{K}((T - \lambda I_V)^m)$. As an element of V, we can write \mathbf{v} as a linear combination of the basis elements of B, or more compactly, there is a vector $\mathbf{u} \in U$ and a scalar α such that $\mathbf{v} = \mathbf{u} + \alpha \mathbf{v}_m$. Then,

$$\begin{aligned} \alpha \left(\lambda_{m} - \lambda\right)^{m} \mathbf{v}_{m} &= \alpha \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{v}_{m}\right) & \text{Theorem EOMP [1006]} \\ &= \mathbf{0} + \alpha \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{v}_{m}\right) & \text{Property Z [660]} \\ &= - \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u}\right) + \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u}\right) + \alpha \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{v}_{m}\right) & \text{Property AI [660]} \\ &= - \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u}\right) + \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u} + \alpha \mathbf{v}_{m}\right) & \text{Theorem LTLC [1093]} \\ &= - \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u}\right) + \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{v}\right) & \text{Theorem LTLC [1093]} \\ &= - \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u}\right) + \mathbf{0} & \text{Definition KLT [1138]} \\ &= - \left(T - \lambda I_{V}\right)^{m} \left(\mathbf{u}\right) & \text{Property Z [660]} \end{aligned}$$

The final expression in this string of equalities is an element of U since U is invariant relative to both T and I_V . The expression at the beginning is a scalar multiple of \mathbf{v}_m , and as such cannot be a nonzero element of U without violating the linear independence of B. So

$$\alpha \left(\lambda_m - \lambda\right)^m \mathbf{v}_m = \mathbf{0}$$

The vector \mathbf{v}_m is nonzero since *B* is linearly independent, so Theorem SMEZV [679] tells us that $\alpha (\lambda_m - \lambda)^m = 0$. From the properties of scalar multiplication, we are confronted with two possibilities.

Our first case is that $\lambda \neq \lambda_m$. Notice then that λ occurs the same number of times along the diagonal in the representations of $T|_U$ and T. Now $\alpha = 0$ and $\mathbf{v} = \mathbf{u} + 0\mathbf{v}_m = \mathbf{u}$. Since \mathbf{v} was chosen as an arbitrary element of $\mathcal{K}((T - \lambda I_V)^m)$, Definition SSET [1603] says that $\mathcal{K}((T - \lambda I_V)^m) \subseteq U$. It is always the case that $\mathcal{K}((T|_U - \lambda I_U)^m) \subseteq \mathcal{K}((T - \lambda I_V)^m)$. However, we can also see that in this case, the opposite set inclusion is true as well. By Definition SE [1604] we have $\mathcal{K}((T|_U - \lambda I_U)^m) = \mathcal{K}((T - \lambda I_V)^m)$. Then

$$\begin{aligned}
\#_{T}(\lambda) &= \#_{T|_{U}}(\lambda) \\
&\geq \dim \left(\mathcal{G}_{T|_{U}}(\lambda)\right) & \text{Induction Hypothesis} \\
&= \dim \left(\mathcal{K}((T|_{U} - \lambda I_{U})^{m-1}\right)) & \text{Theorem GEK [1487]} \\
&= \dim \left(\mathcal{K}((T - \lambda I_{V})^{m})\right) & \text{Theorem KPLT [1450]} \\
&= \dim \left(\mathcal{G}_{T}(\lambda)\right) & \text{Theorem GEK [1487]}
\end{aligned}$$

The second case is that $\lambda = \lambda_m$. Notice then that λ occurs one more time along the diagonal

in the representation of T compared to the representation of $T|_U$. Then

$$(T|_{U} - \lambda I_{U})^{m} (\mathbf{u}) = (T - \lambda I_{V})^{m} (\mathbf{u})$$

$$= (T - \lambda I_{V})^{m} (\mathbf{u}) + \mathbf{0}$$
Property Z [660]
$$= (T - \lambda I_{V})^{m} (\mathbf{u}) + \alpha (\lambda_{m} - \lambda)^{m} \mathbf{v}_{m}$$
Theorem ZSSM [676]
$$= (T - \lambda I_{V})^{m} (\mathbf{u}) + \alpha (T - \lambda I_{V})^{m} (\mathbf{v}_{m})$$
Theorem EOMP [1006]
$$= (T - \lambda I_{V})^{m} (\mathbf{u} + \alpha \mathbf{v}_{m})$$
Theorem LTLC [1093]
$$= (T - \lambda I_{V})^{m} (\mathbf{v})$$

$$= \mathbf{0}$$
Definition KLT [1138]

So $\mathbf{u} \in \mathcal{K}((T|_U - \lambda I_U)^m)$. The vector \mathbf{v} was chosen as an arbitrary member of $\mathcal{K}((T - \lambda I_V)^m)$. From the expression $\mathbf{v} = \mathbf{u} + \alpha \mathbf{v}_m$ we can now see \mathbf{v} also as an element of $\mathcal{K}((T|_U - \lambda I_U)^m)$ plus a scalar multiple of \mathbf{v}_m . This observation yields

$$\dim \left(\mathcal{K}(\left(T - \lambda I_V\right)^m) \right) \le \dim \left(\mathcal{K}(\left(T|_U - \lambda I_U\right)^m) \right) + 1$$

Now count eigenvalues on the diagonal,

$$\#_{T}(\lambda) = \#_{T|_{U}}(\lambda) + 1$$

$$\geq \dim \left(\mathcal{G}_{T|_{U}}(\lambda)\right) + 1$$

$$= \dim \left(\mathcal{K}\left(\left(T|_{U} - \lambda I_{U}\right)^{m-1}\right)\right) + 1$$

$$= \dim \left(\mathcal{K}\left(\left(T|_{U} - \lambda I_{U}\right)^{m}\right)\right) + 1$$

Induction Hypothesis Theorem GEK [1487] Theorem KPLT [1450]

$$\geq \dim \left(\mathcal{K}((T - \lambda I_V)^m) \right)$$

= dim ($\mathcal{G}_T(\lambda)$) Theorem GEK [1487]

In Theorem UTMR [1419] we constructed an upper triangular matrix representation of T where each eigenvalue occurred $\alpha_T(\lambda)$ times on the diagonal. So

 $\alpha_T (\lambda_i) = \#_T(\lambda_i)$ Theorem UTMR [1419] $\geq \dim (\mathcal{G}_T (\lambda_i))$ $\geq \dim (U_i)$ Theorem PSSD [857] $= \alpha_T (\lambda_i)$ Theorem PSSD [857]

Thus, dim $(\mathcal{G}_T(\lambda_i)) = \alpha_T(\lambda_i)$ and by Theorem EDYES [858], $U_i = \mathcal{G}_T(\lambda_i)$ and we can write

$$V = U_1 \oplus U_2 \oplus U_3 \oplus \dots \oplus U_m$$

= $\mathcal{G}_T(\lambda_1) \oplus \mathcal{G}_T(\lambda_2) \oplus \mathcal{G}_T(\lambda_3) \oplus \dots \oplus \mathcal{G}_T(\lambda_m)$

Besides a nice decomposition into invariant subspaces, this proof has a bonus for us.

Theorem DGES Dimension of Generalized Eigenspaces

Suppose $T: V \to V$ is a linear transformation with eigenvalue λ . Then the dimension of the

generalized eigenspace for λ is the algebraic multiplicity of λ , dim $(\mathcal{G}_T(\lambda_i)) = \alpha_T(\lambda_i)$. **Proof** At the very end of the proof of Theorem GESD [1518] we obtain the inequalities

 $\alpha_T(\lambda_i) \leq \dim \left(\mathcal{G}_T(\lambda_i)\right) \leq \alpha_T(\lambda_i)$

which establishes the desired equality.

Subsection JCF Jordan Canonical Form

Now we are in a position to define what we (and others) regard as an especially nice matrix representation. The word "canonical" has at its root, the word "canon," which has various meanings. One is the set of laws established by a church council. Another is a set of writings that are authentic, important or representative. Here we take it to mean the accepted, or best, representative among a variety of choices. Every linear transformation admits a variety of representations, and we will declare one as the best. Hopefully you will agree.

Definition JCF Jordan Canonical Form

A square matrix is in **Jordan canonical form** if it meets the following requirements:

1. The matrix is block diagonal.

- 2. Each block is a Jordan block.
- 3. If $\rho < \lambda$ then the block $J_k(\rho)$ occupies rows with indices greater than the indices of the rows occupied by $J_\ell(\lambda)$.
- 4. If $\rho = \lambda$ and $\ell < k$, then the block $J_{\ell}(\lambda)$ occupies rows with indices greater than the indices of the rows occupied by $J_k(\lambda)$.

Theorem JCFLT

Jordan Canonical Form for a Linear Transformation

Suppose $T: V \to V$ is a linear transformation. Then there is a basis B for V such that the matrix representation of T with the following properties:

- 1. The matrix representation is in Jordan canonical form.
- 2. If $J_k(\lambda)$ is one of the Jordan blocks, then λ is an eigenvalue of T.
- 3. For a fixed value of λ , the largest block of the form $J_k(\lambda)$ has size equal to the index of λ , $\iota_T(\lambda)$.
- 4. For a fixed value of λ , the number of blocks of the form $J_k(\lambda)$ is the geometric multiplicity of λ , $\gamma_T(\lambda)$.

 \triangle

5. For a fixed value of λ , the number of rows occupied by blocks of the form $J_k(\lambda)$ is the algebraic multiplicity of λ , $\alpha_T(\lambda)$.

Proof This theorem is really just the consequence of applying to T, consecutively Theorem GESD [1518], Theorem MRRGE [1514] and Theorem CFNLT [1458].

Theorem GESD [1518] gives us a decomposition of V into generalized eigenspaces, one for each distinct eigenvalue. Since these generalized eigenspaces ar invariant relative to T, this provides a block diagonal matrix representation where each block is the matrix representation of the restriction of T to the generalized eigenspace.

Restricting T to a generalized eigenspace results in a "nearly nilpotent" linear transformation, as stated more precisely in Theorem RGEN [1507]. We unravel Theorem RGEN [1507] in the proof of Theorem MRRGE [1514] so that we can apply Theorem CFNLT [1458] about representations of nilpotent linear transformations.

We know the dimension of a generalized eigenspace is the algebraic multiplicity of the eigenvalue (Theorem DGES [1531]), so the blocks associated with the generalized eigenspaces are square with a size equal to the algebraic multiplicity. In refining the basis for this block, and producing Jordan blocks the results of Theorem CFNLT [1458] apply. The total number of blocks will be the nullity of $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$, which is the geometric multiplicity of λ as an eigenvalue of T (Definition GME [969]). The largest of the Jordan blocks will have size equal to the index of the nilpotent linear transformation $T|_{\mathcal{G}_T(\lambda)} - \lambda I_{\mathcal{G}_T(\lambda)}$, which is exactly the definition of the index of the eigenvalue λ (Definition IE [1508]).

Before we do some examples of this result, notice how close Jordan canonical form is

 \square

to a diagonal matrix. Or, equivalently, notice how close we have come to diagonalizing a matrix (Definition DZM [1037]). We have a matrix representation which has diagonal entries that are the eigenvalues of a matrix. Each occurs on the diagonal as many times as the algebraic multiplicity. However, when the geometric multiplicity is strictly less than the algebraic multiplicity, we have some entries in the representation just above the diagonal (the "superdiagonal"). Furthermore, we have some idea how often this happens if we know the geometric multiplicity and the index of the eigenvalue.

We now recognize just how simple a diagonalizable linear transformation really is. For each eigenvalue, the generalized eigenspace is just the regular eigenspace, and it decomposes into a direct sum of one-dimensional subspaces, each spanned by a different eigenvector chosen from a basis of eigenvectors for the eigenspace.

Some authors create matrix representations of nilpotent linear transformations where the Jordan block has the ones just below the diagonal (the "subdiagonal"). No matter, it is really the same, just different. We have also defined Jordan canonical form to place blocks for the larger eigenvalues earlier, and for blocks with the same eigenvalue, we place the bigger ones earlier. This is fairly standard, but there is no reason we couldn't order the blocks differently. It'd be the same, just different. The reason for choosing *some* ordering is to be assured that there is just *one* canonical matrix representation for each linear transformation.

Example JCF10 Jordan canonical form, size 10 Suppose that $T: \mathbb{C}^{10} \to \mathbb{C}^{10}$ is the linear transformation defined by $T(\mathbf{x}) = A\mathbf{x}$ where

$$A = \begin{bmatrix} -6 & 9 & -7 & -5 & 5 & 12 & -22 & 14 & 8 & 21 \\ -3 & 5 & -3 & -1 & 2 & 7 & -12 & 9 & 1 & 12 \\ 8 & -9 & 8 & 6 & 0 & -14 & 25 & -13 & -4 & -26 \\ -7 & 9 & -7 & -5 & 0 & 13 & -23 & 13 & 2 & 24 \\ 0 & -1 & 0 & -1 & -3 & -2 & 3 & -4 & -2 & -3 \\ 3 & 2 & 1 & 2 & 9 & -1 & 1 & 5 & 5 & -5 \\ -1 & 3 & -3 & -2 & 4 & 3 & -6 & 4 & 4 & 3 \\ 3 & -4 & 3 & 2 & 1 & -5 & 9 & -5 & 1 & -9 \\ 0 & 2 & 0 & 0 & 2 & 2 & -4 & 4 & 2 & 4 \\ -4 & 4 & -5 & -4 & -1 & 6 & -11 & 4 & 1 & 10 \end{bmatrix}$$

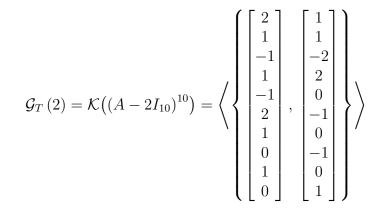
We'll find a basis for \mathbb{C}^{10} that will yield a matrix representation of T in Jordan canonical form. First we find the eigenvalues, and their multiplicities, with the techniques of Chapter E [944].

$\lambda = 2$	$\alpha_T\left(2\right) = 2$	$\gamma_T(2) = 2$
$\lambda = 0$	$\alpha_T\left(0\right) = 3$	$\gamma_T\left(-1\right) = 2$
$\lambda = -1$	$\alpha_T \left(-1 \right) = 5$	$\gamma_T\left(-1\right) = 2$

For each eigenvalue, we can compute a generalized eigenspace. By Theorem GESD [1518] we know that \mathbb{C}^{10} will decompose into a direct sum of these eigenspaces, and we can restrict T to each part of this decomposition. At this stage we know that the Jordan canonical

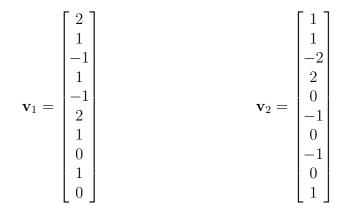
form will be block diagonal with blocks of size 2, 3 and 5, since the dimensions of the generalized eigenspaces are equal to the algebraic multiplicities of the eigenvalues (Theorem DGES [1531]). The geometric multiplicities tell us how many Jordan blocks occupy each of the three larger blocks, but we will discuss this as we analyze each eigenvalue. We do not yet know the index of each eigenvalue (though we can easily infer it for $\lambda = 2$) and even if we did have this information, it only determines the size of the largest Jordan block (per eigenvalue). We will press ahead, considering each eigenvalue one at a time.

The eigenvalue $\lambda = 2$ has "full" geometric multiplicity, and is not an impediment to diagonalizing T. We will treat it in full generality anyway. First we compute the generalized eigenspace. Since Theorem GEK [1487] says that $\mathcal{G}_T(2) = \mathcal{K}((T - 2I_{\mathbb{C}^{10}})^{10})$ we can compute this generalized eigenspace as a null space derived from the matrix A,



The restriction of T to $\mathcal{G}_T(2)$ relative to the two basis vectors above has a matrix representation that is a 2 × 2 diagonal matrix with the eigenvalue $\lambda = 2$ as the diagonal entries. So

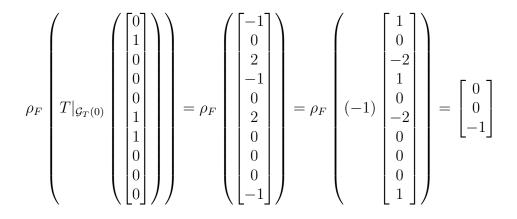
these two vectors will be the first two vectors in our basis for \mathbb{C}^{10} ,

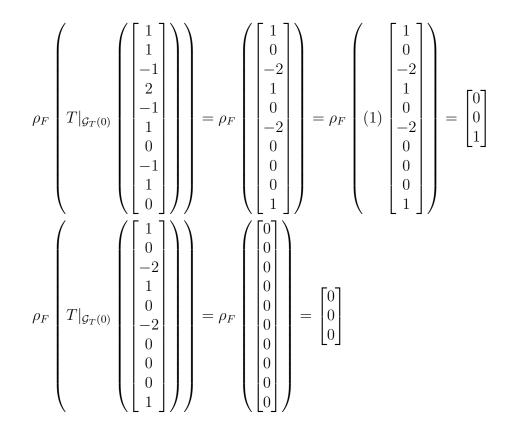


Notice that it was not strictly necessary to compute the 10-th power of $A - 2I_{10}$. With $\alpha_T(2) = \gamma_T(2)$ the null space of the matrix $A - 2I_{10}$ contains all of the generalized eigenvectors of T for the eigenvalue $\lambda = 2$. But there was no harm in computing the 10-th power either. This discussion is equivalent to the observation that the linear transformation $T|_{\mathcal{G}_T(2)}: \mathcal{G}_T(2) \to \mathcal{G}_T(2)$ is nilpotent of index 1. In other words, $\iota_T(2) = 1$.

The eigenvalue $\lambda = 0$ will not be quite as simple, since the geometric multiplicity is strictly less than the geometric multiplicity. As before, we first compute the generalized eigenspace. Since Theorem GEK [1487] says that $\mathcal{G}_T(0) = \mathcal{K}((T - 0I_{\mathbb{C}^{10}})^{10})$ we can compute this generalized eigenspace as a null space derived from the matrix A,

So dim $(\mathcal{G}_T(0)) = 3 = \alpha_T(0)$, as expected. We will use these three basis vectors for the generalized eigenspace to construct a matrix representation of $T|_{\mathcal{G}_T(0)}$, where F is being defined implicitly as the basis of $\mathcal{G}_T(0)$. We construct this representation as usual, applying Definition MR [1286],





So we have the matrix representation

J

$$M = M_{F,F}^{T|_{\mathcal{G}_T(0)}} = \begin{bmatrix} 0 & 0 & 0\\ 0 & 0 & 0\\ -1 & 1 & 0 \end{bmatrix}$$

By Theorem RGEN [1507] we can obtain a nilpotent matrix from this matrix representation by subtracting the eigenvalue from the diagonal elements, and then we can apply Theorem CFNLT [1458] to $M - (0)I_3$. First check that $(M - (0)I_3)^2 = \mathcal{O}$, so we know that the index of $M - (0)I_3$ as a nilpotent matrix, and that therefore $\lambda = 0$ is an eigenvalue of T with index $2, \iota_T(0) = 2$. To determine a basis of \mathbb{C}^3 that converts $M - (0)I_3$ to canonical form, we need the null spaces of the powers of $M - (0)I_3$. For convenience, set $N = M - (0)I_3$.

$$\mathcal{N}(N^{1}) = \left\langle \left\{ \begin{bmatrix} 1\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle$$
$$\mathcal{N}(N^{2}) = \left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix} \right\} \right\rangle = \mathbb{C}^{3}$$

Then we choose a vector from $\mathcal{N}(N^2)$ that is not an element of $\mathcal{N}(N^1)$. Any vector with unequal first two entries will fit the bill, say

$$\mathbf{z}_{2,1} = \begin{bmatrix} 1\\0\\0 \end{bmatrix}$$

where we are employing the notation in Theorem CFNLT [1458]. The next step is to multiply this vector by N to get part of the basis for $\mathcal{N}(N^1)$,

$$\mathbf{z}_{1,1} = N\mathbf{z}_{2,1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

We need a vector to pair with $\mathbf{z}_{1,1}$ that will make a basis for the two-dimensional subspace $\mathcal{N}(N^1)$. Examining the basis for $\mathcal{N}(N^1)$ we see that a vector with its first two entries equal will do the job.

$$\mathbf{z}_{1,2} = \begin{bmatrix} 1\\1\\0 \end{bmatrix}$$

Reordering, we find the basis,

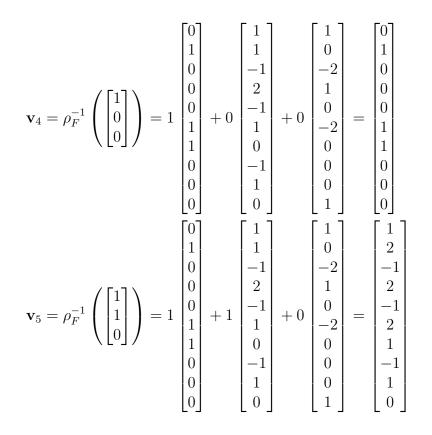
$$C = \{\mathbf{z}_{1,1}, \, \mathbf{z}_{2,1}, \, \mathbf{z}_{1,2}\} = \left\{ \begin{bmatrix} 0\\0\\-1 \end{bmatrix}, \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\0 \end{bmatrix} \right\}$$

From this basis, we can get a matrix representation of N (when viewed as a linear transformation) relative to the basis C for \mathbb{C}^3 ,

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} J_2(0) & \mathcal{O} \\ \mathcal{O} & J_1(0) \end{bmatrix}$$

Now we add back the eigenvalue $\lambda = 0$ to the representation of N to obtain a representation for M. Of course, with an eigenvalue of zero, the change is not apparent, so we won't display the same matrix again. This is the second block of the Jordan canonical form for T. However, the three vectors in C will not suffice as basis vectors for the domain of T — they have the wrong size! The vectors in C are vectors in the domain of a linear transformation defined by the matrix M. But M was a matrix representation of $T|_{\mathcal{G}_T(0)} - 0I_{\mathcal{G}_T(0)}$ relative to the basis F for $\mathcal{G}_T(0)$. We need to "uncoordinatize" each of the basis vectors in C to produce a linear combination of vectors in F that will be an element of the generalized eigenspace $\mathcal{G}_T(0)$. These will be the next three vectors of our final answer, a basis for \mathbb{C}^{10} that has a pleasing matrix representation.

$$\mathbf{v}_{3} = \rho_{F}^{-1} \left(\begin{bmatrix} 0\\0\\-1 \end{bmatrix} \right) = 0 \begin{bmatrix} 0\\1\\0\\0\\0\\-1 \end{bmatrix} + 0 \begin{bmatrix} 1\\1\\-1\\2\\-1\\1\\0\\-1\\1\\0\\0 \end{bmatrix} + (-1) \begin{bmatrix} 1\\0\\-2\\1\\0\\-2\\1\\0\\-1\\1\\0\\0\end{bmatrix} = \begin{bmatrix} -1\\0\\2\\-1\\0\\0\\0\\0\\0\\-1\end{bmatrix}$$

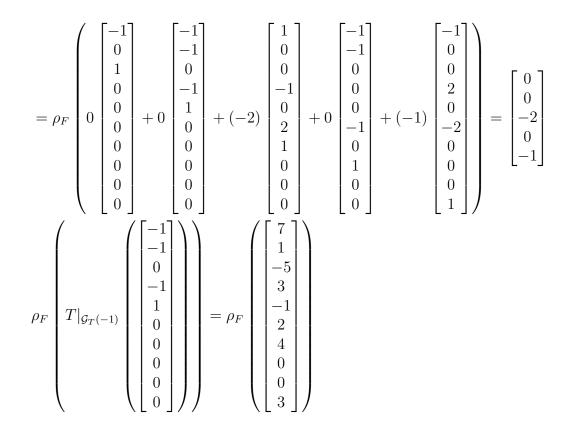


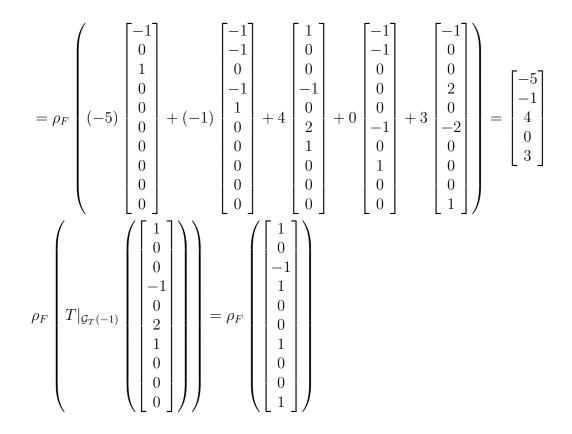
Five down, five to go. Basis vectors, that is. $\lambda = -1$ is the smallest eigenvalue, but it will require the most computation. First we compute the generalized eigenspace. Since Theorem

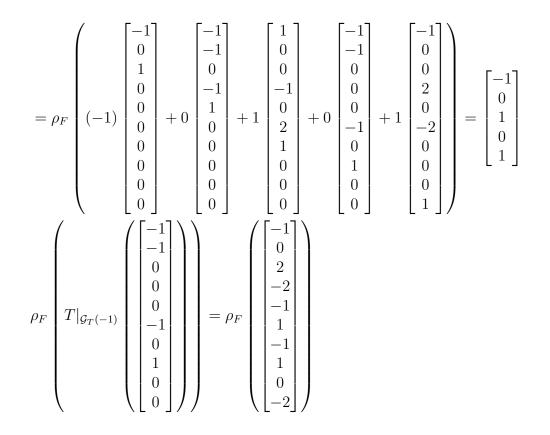
GEK [1487] says that $\mathcal{G}_T(-1) = \mathcal{K}((T - (-1)I_{\mathbb{C}^{10}})^{10})$ we can compute this generalized eigenspace as a null space derived from the matrix A,

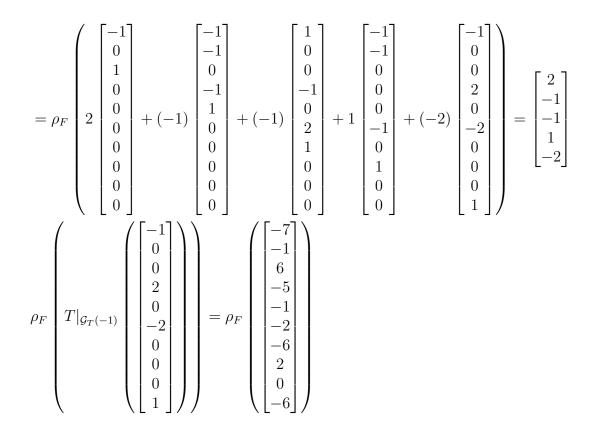
So dim $(\mathcal{G}_T(-1)) = 5 = \alpha_T(-1)$, as expected. We will use these five basis vectors for the generalized eigenspace to construct a matrix representation of $T|_{\mathcal{G}_T(-1)}$, where F is being recycled and defined now implicitly as the basis of $\mathcal{G}_T(-1)$. We construct this representation as usual, applying Definition MR [1286],

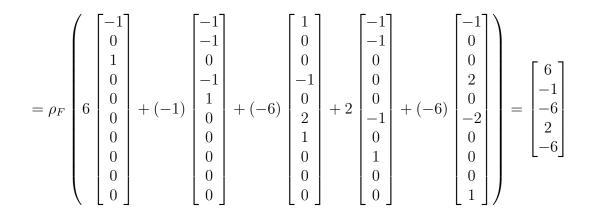
$$\rho_F \left(T|_{\mathcal{G}_T(-1)} \left(\begin{bmatrix} -1\\ 0\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} \right) \right) = \rho_F \left(\begin{bmatrix} -1\\ 0\\ 0\\ 0\\ 0\\ -2\\ -2\\ 0\\ 0\\ 0\\ -1 \end{bmatrix} \right)$$











So we have the matrix representation of the restriction of T (again recycling and redefining the matrix M)

$$M = M_{F,F}^{T|_{\mathcal{G}_T(-1)}} = \begin{bmatrix} 0 & -5 & -1 & 2 & 6\\ 0 & -1 & 0 & -1 & -1\\ -2 & 4 & 1 & -1 & -6\\ 0 & 0 & 0 & 1 & 2\\ -1 & 3 & 1 & -2 & -6 \end{bmatrix}$$

By Theorem RGEN [1507] we can obtain a nilpotent matrix from this matrix representation by subtracting the eigenvalue from the diagonal elements, and then we can apply Theorem CFNLT [1458] to $M - (-1)I_5$. First check that $(M - (-1)I_5)^3 = \mathcal{O}$, so we know that the index of $M - (-1)I_5$ as a nilpotent matrix, and that therefore $\lambda = -1$ is an eigenvalue of T with index 3, $\iota_T(-1) = 3$. To determine a basis of \mathbb{C}^5 that converts $M - (-1)I_5$ to canonical form, we need the null spaces of the powers of $M - (-1)I_5$. Again, for convenience, set $N = M - (-1)I_5$.

Then we choose a vector from $\mathcal{N}(N^3)$ that is not an element of $\mathcal{N}(N^2)$. The sum of the four basis vectors for $\mathcal{N}(N^2)$ sum to a vector with all five entries equal to 1. We will mess with

the first entry to create a vector not in $\mathcal{N}(N^2)$,

$$\mathbf{z}_{3,1} = \begin{bmatrix} 0\\1\\1\\1\\1\\1\end{bmatrix}$$

where we are employing the notation in Theorem CFNLT [1458]. The next step is to multiply this vector by N to get a portion of the basis for $\mathcal{N}(N^2)$,

$$\mathbf{z}_{2,1} = N\mathbf{z}_{3,1} = \begin{bmatrix} 1 & -5 & -1 & 2 & 6 \\ 0 & 0 & 0 & -1 & -1 \\ -2 & 4 & 2 & -1 & -6 \\ 0 & 0 & 0 & 2 & 2 \\ -1 & 3 & 1 & -2 & -5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ -1 \\ 4 \\ -3 \end{bmatrix}$$

We have a basis for the two-dimensional subspace $\mathcal{N}(N^1)$ and we can add to that the vector $\mathbf{z}_{2,1}$ and we have three of four basis vectors for $\mathcal{N}(N^2)$. These three vectors span the subspace we call Q_2 . We need a fourth vector outside of Q_2 to complete a basis of the four-dimensional

subspace $\mathcal{N}(N^2)$. Check that the vector

$$\mathbf{z}_{2,2} = \begin{bmatrix} 3\\1\\3\\1\\1\\1\end{bmatrix}$$

is an element of $\mathcal{N}(N^2)$ that lies outside of the subspace Q_2 . This vector was constructed by getting a nice basis for Q_2 and forming a linear combination of this basis that specifies three of the five entries of the result. Of the remaining two entries, one was changed to move the vector outside of Q_2 and this was followed by a change to the remaining entry to place the vector into $\mathcal{N}(N^2)$. The vector $\mathbf{z}_{2,2}$ is the lone basis vector for the subspace we call R_2 .

The remaining two basis vectors are easy to come by. They are the result of applying N to each of the two most recently determined basis vectors,

$$\mathbf{z}_{1,1} = N\mathbf{z}_{2,1} = \begin{bmatrix} 3\\ -1\\ 0\\ 2\\ -2 \end{bmatrix} \qquad \qquad \mathbf{z}_{1,2} = N\mathbf{z}_{2,2} = \begin{bmatrix} 3\\ -2\\ -3\\ 4\\ -4 \end{bmatrix}$$

Now we reorder these basis vectors, to arrive at the basis

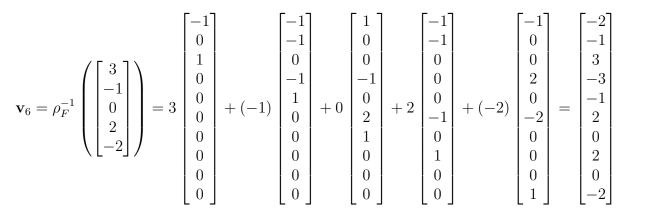
$$C = \{\mathbf{z}_{1,1}, \, \mathbf{z}_{2,1}, \, \mathbf{z}_{3,1}, \, \mathbf{z}_{1,2}, \, \mathbf{z}_{2,2}\} = \left\{ \begin{bmatrix} 3\\-1\\0\\2\\-2 \end{bmatrix}, \begin{bmatrix} 2\\-2\\-1\\4\\-3 \end{bmatrix}, \begin{bmatrix} 0\\1\\1\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} 3\\-2\\-3\\4\\-4 \end{bmatrix}, \begin{bmatrix} 3\\1\\3\\1\\1\\1 \end{bmatrix} \right\}$$

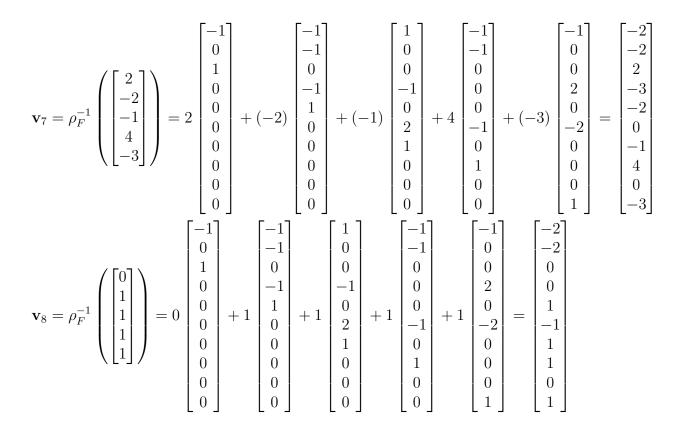
A matrix representation of N relative to C is

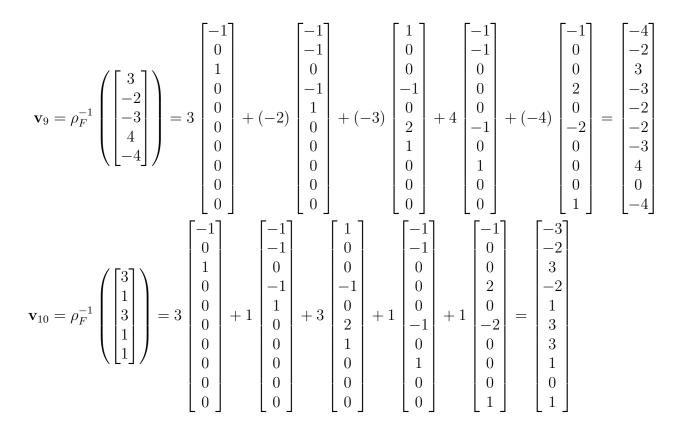
To obtain a matrix representation of M, we add back in the matrix $(-1)I_5$, placing the eigenvalue back along the diagonal, and slightly modifying the Jordan blocks,

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} J_3(-1) & \mathcal{O} \\ \mathcal{O} & J_2(-1) \end{bmatrix}$$

The basis C yields a pleasant matrix representation for the *restriction* of the linear transformation T - (-1)I to the generalized eigenspace $\mathcal{G}_T(-1)$. However, we must remember that these vectors in \mathbb{C}^5 are representations of vectors in \mathbb{C}^{10} relative to the basis F. Each needs to be "un-coordinatized" before joining our final basis. Here we go,







To summarize, we list the entire basis $B = {\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_{10}},$

The resulting matrix representation is

If you are not inclined to check all of these computations, here are a few that should convince you of the amazing properties of the basis B. Compute the matrix-vector products $A\mathbf{v}_i$, $1 \leq i \leq 10$. In each case the result will be a vector of the form $\lambda \mathbf{v}_i + \delta \mathbf{v}_{i-1}$, where λ is one of the eigenvalues (you should be able to predict ahead of time *which* one) and $\delta \in \{0, 1\}$.

Alternatively, if we can write inputs to the linear transformation T as linear combinations of the vectors in B (which we can do uniquely since B is a basis, Theorem VRRB [755]), then the "action" of T is reduced to a matrix-vector product with the exceedingly simple matrix that is the Jordan canonical form. Wow!

Subsection CHT Cayley-Hamilton Theorem

Jordan was a French mathematician who was active in the late 1800's. Cayley and Hamilton were 19th-century contemporaries of Jordan from Britain. The theorem that bears their names is perhaps one of the most celebrated in basic linear algebra. While our result applies only to vector spaces and linear transformations with scalars from the set of complex numbers, \mathbb{C} , the result is equally true if we restrict our scalars to the real numbers, \mathbb{R} . It says that every matrix satisfies its own characteristic polynomial.

Theorem CHT Cayley-Hamilton Theorem

Suppose A is a square matrix with characteristic polynomial $p_A(x)$. Then $p_A(A) = \mathcal{O}$. \Box

Proof Suppose *B* and *C* are similar matrices via the matrix *S*, $B = S^{-1}CS$, and q(x) is any polynomial. Then q(B) is similar to q(C) via *S*, $q(B) = S^{-1}q(C)S$. (See Example HPDM [1051] for hints on how to convince yourself of this.)

By Theorem JCFLT [1533] and Theorem SCB [1377] we know A is similar to a matrix, J, in Jordan canonical form. Suppose $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_m$ are the distinct eigenvalues of A (and are therefore the eigenvalues and diagonal entries of J). Then by Theorem EMRCP [962] and Definition AME [968], we can factor the characteristic polynomial as

$$p_A(x) = (x - \lambda_1)^{\alpha_A(\lambda_1)} (x - \lambda_2)^{\alpha_A(\lambda_2)} (x - \lambda_3)^{\alpha_A(\lambda_3)} \cdots (x - \lambda_m)^{\alpha_A(\lambda_m)}$$

On substituting the matrix J we have

$$p_A(J) = (J - \lambda_1 I)^{\alpha_A(\lambda_1)} (J - \lambda_2 I)^{\alpha_A(\lambda_2)} (J - \lambda_3 I)^{\alpha_A(\lambda_3)} \cdots (J - \lambda_m I)^{\alpha_A(\lambda_m)}$$

The matrix $J - \lambda_k I$ will be block diagonal, and the block arising from the generalized eigenspace for λ_k will have zeros along the diagonal. Suitably adjusted for matrices (rather than linear transformations), Theorem RGEN [1507] tells us this matrix is nilpotent. Since the size of this nilpotent matrix is equal to the algebraic multiplicity of λ_k , the power $(J - \lambda_k I)^{\alpha_A(\lambda_k)}$ will be a zero matrix (Theorem KPNLT [1453]) in the location of this block.

Repeating this argument for each of the m eigenvalues will place a zero block in some term of the product at every location on the diagonal. The entire product will then be zero blocks on the diagonal, and zero off the diagonal. In other words, it will be the zero matrix. Since A and J are similar, $p_A(A) = p_A(J) = \mathcal{O}$.

Appendix CN Computation Notes

Section MMA Mathematica

Computation Note ME.MMA Matrix Entry

Matrices are input as lists of lists, since a list is a basic data structure in *Mathematica*. A matrix is a list of rows, with each row entered as a list. *Mathematica* uses braces $((\{,,\}))$ to delimit lists. So the input

$$a = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10, 11, 12\}\}$$

would create a 3×4 matrix named **a** that is equal to

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

To display a matrix named **a** "nicely" in *Mathematica*, type MatrixForm[a], and the output will be displayed with rows and columns. If you just type **a**, then you will get a list of lists, like how you input the matrix in the first place.

Computation Note RR.MMA Row Reduce

If a is the name of a matrix in Mathematica, then the command RowReduce[a] will output the reduced row-echelon form of the matrix.

Computation Note LS.MMA Linear Solve

Mathematica will solve a linear system of equations using the LinearSolve[] command. The inputs are a matrix with the coefficients of the variables (but not the column of constants), and a list containing the constant terms of each equation. This will look a bit odd, since the lists in the matrix are rows, but the column of constants is also input as a list and so looks like a row rather than a column. The result will be a single solution (even if there are infinitely many), reported as a list, or the statement that there is no solution. When there are infinitely many, the single solution reported is exactly that solution used in the proof of Theorem RCLS [120], where the free variables are all set to zero, and the dependent variables come along with values from the final column of the row-reduced matrix.

As an example, Archetype A [1642] is

$$x_1 - x_2 + 2x_3 = 1$$

$$2x_1 + x_2 + x_3 = 8$$

$$x_1 + x_2 = 5$$

To ask Mathematica for a solution, enter

```
\texttt{LinearSolve}[ \ \{\{1, \ -1, \ 2\}, \{2, \ 1, \ 1\}, \{1, \ 1, \ 0\}\}, \ \{1, \ 8, \ 5\} \ ]
```

and you will get back the single solution

 $\{3, 2, 0\}$

We will see later how to coax *Mathematica* into giving us infinitely many solutions for this system (Computation VFSS.MMA [1571]).

Computation Note VLC.MMA Vector Linear Combinations

Contributed by Robert Beezer Vectors in *Mathematica* are represented as lists, written and displayed horizontally. For example, the vector

$$\mathbf{v} = \begin{bmatrix} 1\\2\\3\\4 \end{bmatrix}$$

- -

would be entered and named via the command

$$\mathtt{v} = \{\mathtt{1},\,\mathtt{2},\,\mathtt{3},\,\mathtt{4}\}$$

Vector addition and scalar multiplication are then very natural. If u and v are two lists of equal length, then

$$2u + (-3)v$$

will compute the correct vector and return it as a list. If u and v have different sizes, then *Mathematica* will complain about "objects of unequal length."

Computation Note NS.MMA Null Space

Given a matrix A, Mathematica will compute a set of column vectors whose span is the null space of the matrix with the NullSpace[] command. Perhaps not coincidentally, this set

is exactly $\{\mathbf{z}_j | 1 \leq j \leq n - r\}$. However, Mathematica prefers to output the vectors in the opposite order than one we have chosen. Here's a small example.

Begin with the 3×4 matrix A, and its row-reduced version B,

$$A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 3 & 4 & 1 & -2 \\ -1 & 1 & -5 & 3 \end{bmatrix} \xrightarrow{\text{RREF}} B = \begin{bmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We could extract entries from B to build the vectors \mathbf{z}_1 and \mathbf{z}_2 according to Theorem SSNS [283] and describe $\mathcal{N}(A)$ as a span of the set $\{\mathbf{z}_1, \mathbf{z}_2\}$. Instead, if a has been set to A, then executing the command NullSpace[a] yields the list of lists (column vectors),

$$\{\{2, -1, 0, 1\}, \{-3, 2, 1, 0\}\}$$

Notice how our z_1 is second in the list. To "correct" this we can use a list-processing command from Mathematica, Reverse[], as follows,

Reverse[NullSpace[a]]

and receive the output in our preferred order. Give it a try yourself.

Computation Note VFSS.MMA Vector Form of Solution Set

Suppose that A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{C}^m$ is a column vector. We might wish to find all of the solutions to the linear system $\mathcal{LS}(A, \mathbf{b})$. Mathematica's LinearSolve[A, b] will return at most one solution (Computation LS.MMA [1567]). However, when the system is consistent, then this one solution reported is exactly the vector \mathbf{c} , described in the statement of Theorem VFSLS [244].

The vectors \mathbf{u}_j , $1 \leq j \leq n-r$ of Theorem VFSLS [244] are exactly the output of Mathematica's NullSpace[] command, though Mathematica lists them in the opposite order from the order we have chosen. These are the same vectors listed as \mathbf{z}_j , $1 \leq j \leq n-r$ in Theorem SSNS [283]. With c produced from the LinearSolve[] command, and the \mathbf{u}_j coming from the NullSpace[] command we can use Mathematica's symbolic manipulation commands to create an expression that describes all of the solutions.

Begin with the system $\mathcal{LS}(A, \mathbf{b})$. Row-reduce A (Computation RR.MMA [1567]) and identify the free variables by determining the non-pivot columns. Suppose, for the sake of argument, that we have the three free variables x_3 , x_7 and x_8 . Then the following command will build an expression for an arbitrary solution:

LinearSolve[A, b]+{x8, x7, x3}.NullSpace[A]

Be sure to include the "dot" right before the NullSpace[] command — it has the effect of

creating a linear combination of the vectors in the null space, using scalars that are symbols reminiscent of the variables.

A concrete example should help here. Suppose we want a solution set for the linear system with coefficient matrix A and vector of constants \mathbf{b} ,

$$A = \begin{bmatrix} 1 & 2 & 3 & -5 & 1 & -1 & 2 \\ 2 & 4 & 0 & 8 & -4 & 1 & -8 \\ 3 & 6 & 4 & 0 & -2 & 5 & 7 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 8 \\ 1 \\ -5 \end{bmatrix}$$

If we were to apply Theorem VFSLS [244], we would extract the components of \mathbf{c} and \mathbf{u}_j from the row-reduced version of the augmented matrix of the system (obtained with Mathematica, Computation RR.MMA [1567]),

$$\begin{bmatrix} 1 & 2 & 0 & 4 & -2 & 0 & -5 & 2 \\ 0 & 0 & 1 & -3 & 1 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & -3 \end{bmatrix}$$

Instead, we will use this augmented matrix in reduced row-echelon form only to identify the free variables. In this example, we locate the non-pivot columns and see that x_2 , x_4 , x_5 and x_7 are free. If we have set **a** to the coefficient matrix and **b** to the vector of constants, then we execute the Mathematica command,

```
LinearSolve[a, b]+\{x7, x5, x4, x2\}.NullSpace[a]
```

As output we obtain the column vector (list),

$$\begin{bmatrix} 2 - 2 x 2 - 4 x 4 + 2 x 5 + 5 x 7 \\ x 2 \\ 1 + 3 x 4 - x 5 - 3 x 7 \\ x 4 \\ x 5 \\ -3 - 2 x 7 \\ x 7 \end{bmatrix}$$

Computation Note GSP.MMA Gram-Schmidt Procedure

Mathematica has a built-in routine that will do the Gram-Schmidt procedure (Theorem GSP [413]). The input is a set of vectors, which must be linearly independent. This is written as a list, containing lists that are the vectors. Let **a** be such a list of lists, containing the vectors \mathbf{v}_i , $1 \leq i \leq p$ from the statement of the theorem. You will need to first load the right Mathematica package — execute <<LinearAlgebra'Orthogonalization' to make this happen. Then execute GramSchmidt[a]. The output will be another list of lists containing the vectors \mathbf{u}_i , $1 \leq i \leq p$ from the statement of the theorem. Mathematica will complain if you do not provide a linearly independent set as input (try it!).

An example. Suppose our linearly independent set (check this!) is

$$S = \left\{ \begin{bmatrix} -1\\4\\1\\0\\3\end{bmatrix}, \begin{bmatrix} 0\\3\\0\\-3\\-3\end{bmatrix}, \begin{bmatrix} -1\\2\\0\\-1\\-2\\-3\\-1\\-2\end{bmatrix}, \begin{bmatrix} -1\\-2\\-3\\1\\4\end{bmatrix}, \begin{bmatrix} 1\\6\\-1\\4\\6\end{bmatrix} \right\}$$

The output of the GramSchmidt[] command will be the set,

$$T = \left\{ \begin{bmatrix} -\frac{1}{3\sqrt{3}} \\ \frac{4}{3\sqrt{3}} \\ \frac{1}{3\sqrt{3}} \\ \frac{1}{3\sqrt{3}} \\ 0 \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{1}{12\sqrt{15}} \\ \frac{23}{12\sqrt{15}} \\ -\frac{1}{12\sqrt{15}} \\ \frac{3\sqrt{\frac{3}{5}}}{4} \\ -\frac{\sqrt{\frac{3}{5}}}{2} \end{bmatrix}, \begin{bmatrix} -\frac{37}{4\sqrt{685}} \\ \frac{29}{4\sqrt{685}} \\ -\frac{3}{4\sqrt{685}} \\ -\frac{79}{4\sqrt{685}} \\ -\frac{79}{4\sqrt{685}} \\ \frac{5\sqrt{\frac{5}{137}}}{2} \end{bmatrix}, \begin{bmatrix} -\frac{337}{2\sqrt{120423}} \\ -\frac{37}{6\sqrt{120423}} \\ -\frac{37}{6\sqrt{120423}} \\ \frac{337}{6\sqrt{120423}} \\ \frac{337}{6\sqrt{120423}} \\ \frac{50}{\sqrt{120423}} \end{bmatrix}, \begin{bmatrix} \frac{23}{\sqrt{879}} \\ \frac{26}{3\sqrt{879}} \\ -\frac{44}{3\sqrt{879}} \\ -\frac{23}{3\sqrt{879}} \\ \frac{1}{\sqrt{879}} \end{bmatrix} \right\}$$

Ugly, but true. At this stage, you might just as well be encouraged to think of the Gram-Schmidt procedure as a computational black box, linearly independent set in, orthogonal span-preserving set out.

To check that the output set is orthogonal, we can easily check the orthogonality of individual pairs of vectors. Suppose the output was set equal to b (say via b=GramSchmidt[a]). We can extract the individual vectors of c as "parts" with syntax like c[[3]], which would return the third vector in the set. When our vectors have only real number entries,

we can accomplish an innerproduct with a "dot." So, for example, you should discover that c[[3]].c[[5]] will return zero. Try it yourself with another pair of vectors.

Computation Note TM.MMA Transpose of a Matrix

Contributed by Robert Beezer

Suppose a is the name of a matrix stored in *Mathematica*. Then Transpose[a] will create the transpose of a.

Computation Note MM.MMA Matrix Multiplication

If A and B are matrices defined in *Mathematica*, then A.B will return the product of the two matrices (notice the dot between the matrices). If A is a matrix and \mathbf{v} is a vector, then A.v will return the vector that is the matrix-vector product of A and v. In every case the sizes of the matrices and vectors need to be correct.

Some examples:

$$\begin{split} \{\{1,\,2\},\,\{3,\,4\}\}.\{\{5,\,6,\,7\},\,\{8,\,9,\,10\}\} &= \{\{21,\,24,\,27\},\,\{47,\,54,\,61\}\} \\ \{\{1,\,2\},\,\{3,\,4\}\}.\{\{5\},\,\{6\}\} &= \{\{17\},\,\{39\}\} \\ \{\{1,\,2\},\,\{3,\,4\}\}.\{5,\,6\} &= \{17,\,39\} \end{split}$$

Understanding the difference between the last two examples will go a long way to explaining how some *Mathematica* constructs work.

Computation Note MI.MMA Matrix Inverse

If A is a matrix defined in *Mathematica*, then Inverse[A] will return the inverse of A, should it exist. In the case where A does not have an inverse *Mathematica* will tell you the matrix is singular (see Theorem NI [542]).

Section TI86 Texas Instruments 86

Computation Note ME.TI86 Matrix Entry

On the TI-86, press the MATRX key (Yellow-7). Press the second menu key over, F2, to bring up the EDIT screen. Give your matrix a name, one letter or many, then press ENTER. You can then change the size of the matrix (rows, then columns) and begin editing individual entries (which are initially zero). ENTER will move you from entry to entry, or the down arrow key will move you to the next row. A menu gives you extra options for editing.

Matrices may also be entered on the home screen as follows. Use brackets ([,]) to enclose rows with elements separated by commas. Group rows, in order, into a final set of brackets (with no commas between rows). This can then be stored in a name with the STO key. So, for example,

$$[[1,2,3,4] \; [5,6,7,8] \; [9,10,11,12]] \to A$$

will create a matrix named A that is equal to

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix}$$

Computation Note RR.TI86 Row Reduce

If A is the name of a matrix stored in the TI-86, then the command rref A will return the reduced row-echelon form of the matrix. This command can also be found by pressing the MATRX key, then F4 for OPS, and finally, F5 for rref.

Note that this command will not work for a matrix with more rows than columns. (Ed. Not sure just why this is!) A work-around is to pad the matrix with extra columns of zeros until the matrix is square.

Computation Note VLC.TI86 Vector Linear Combinations

Contributed by Robert Beezer

Vector operations on the TI-86 can be accessed via the VECTR key, which is Yellow-8. The EDIT tool appears when the F2 key is pressed. After providing a name and giving a "dimension" (the size) then you can enter the individual entries, one at a time. Vectors can also be entered on the home screen using brackets ([,]). To create the vector

$$\mathbf{v} = \begin{bmatrix} 1\\2\\3\\4 \end{bmatrix}$$

use brackets and the store key (STO),

$$[1,\,2,\,3,\,4]\rightarrow v$$

Vector addition and scalar multiplication are then very natural. If $\,u\,$ and $\,v\,$ are two vectors of equal size, then

$$2 * u + (-3) * v$$

will compute the correct vector and display the result as a vector.

Computation Note TM.TI86 Transpose of a Matrix

Contributed by Eric Fickenscher

Suppose A is the name of a matrix stored in the TI-86. Use the command $A^{\rm T}$ to transpose A. This command can be found by pressing the $MATRX\,$ key, then $F3\,$ for MATH, then $F2\,$ for $^{\rm T}.$

Section TI83 Texas Instruments 83

Computation Note ME.TI83 Matrix Entry

Contributed by Douglas Phelps On the TI-83, press the MATRX key. Press the right arrow key twice so that EDIT is highlighted. Move the cursor down so that it is over the desired letter of the matrix and press ENTER. For example, let's call our matrix B, so press the down arrow once and press ENTER. To enter a 2×3 matrix, press 2 ENTER 3 ENTER. To create the matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

press 1 ENTER 2 ENTER 3 ENTER 4 ENTER 5 ENTER 6 ENTER.

Computation Note RR.TI83 Row Reduce

Contributed by Douglas Phelps

Suppose B is the name of a matrix stored in the TI-83. Press the MATRX key. Press the right arrow key once so that MATH is highlighted. Press the down arrow eleven times so that rref (is highlighted, then press ENTER . to choose the matrix B, press MATRX, then the down arrow once followed by ENTER. Supply a right parenthesis () and press ENTER.

Note that this command will not work for a matrix with more rows than columns. (Ed. Not sure just why this is!) A work-around is to pad the matrix with extra columns of zeros until the matrix is square.

Computation Note VLC.TI83 Vector Linear Combinations

Contributed by Douglas Phelps

Entering a vector on the TI-83 is the same process as entering a matrix. You press 4 ENTER 3 ENTER for a 4×3 matrix. Likewise, you press 4 ENTER 1 ENTER for a vector of size 4. To multiply a vector by 8, press the number 8, then press the MATRX key, then scroll down to the letter you named your vector (A, B, C, etc) and press ENTER.

To add vectors A and B for example, press the MATRX key, then ENTER. Then press the + key. Then press the MATRX key, then the down arrow once, then ENTER. [A] + [B] will appear on the screen. Press ENTER.

Section SAGE SAGE: Open Source Mathematics Software

Computation Note R.SAGE Rings

Contributed by Steve Canfield SAGE uses different rings to denote the type of an object. The rings are as follows:

ZZ: The set of integersQQ: The set of rational numbersRR: The real numbersCC: The complex numbers

Most objects in SAGE will tell you which they are using with the base_ring() command. Keep this in mind, especially when row reducing or factoring. Here's a quick example of where you might go wrong.

```
\begin{split} \mathbf{m} &= \mathtt{matrix}([[2,3],[4,7]]) \\ \mathtt{m.base\_ring}() \\ \mathtt{IntegerRing} \\ \mathtt{m.echelon\_form}() \\ & \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \end{split}
```

As you can clearly see, m isn't even in reduced row-echelon form. This is because m is defined over the ZZ. You have to create matrices with the correct ring or you will get this type of odd result. This problem comes up in more places than just calculating the reduced row-echelon form, so unless you are specifically working with integers take note.

Computation Note ME.SAGE Matrix Entry

Contributed by Steve Canfield

A matrix in SAGE can be made a few ways. The first is simply to define the matrix as an array of rows. SAGE uses brackets ([,]) to delimit arrays. So the input

 $\mathtt{a} = \mathtt{matrix}([[1,2,3,4],[5,6,7,8],[9,10,11,12]])$

would create a 3×4 matrix named **a** that is equal to

1	2	3	4]
5	6	$\overline{7}$	8
9	10	11	12

SAGE will guess what type of matrix you are working with based on the inputs. If all the entries are integers, you will get back an integer matrix. If your matrix contains an entry in the \mathbb{R} or \mathbb{C} space, the matrix will be of those types. This can cause problems as integers cannot become fractions, which is an issue when calculating reduced row-echelon form. We therefore recommend using the following construction to make your matrices,

a = matrix(QQ, [[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]])

This gives you a matrix over the rational numbers which will be sufficient for most of the course. If your matrix has entries that are complex numbers you would replace the $\,QQ\,$ with CC.

To display a matrix named $\tt a$, type $\tt a$, and the output will be displayed with rows and columns. If you type <code>latex(a)</code> you will get ${\it L}\!AT_{E\!}X$ code to display the matrix. Very handy.

Computation Note RR.SAGE Row Reduce

Contributed by Steve Canfield and Robert Beezer

Row-reducing a matrix is a simple operation in SAGE. However, because of Sage's flexibility with different types of numbers (integers, rationals, reals, complexes), we need to be a bit more careful.

If a is a matrix entered in in SAGE (see Computation ME.SAGE [1584]) then a.echelon_form() will return a new matrix that is the reduced row-echelon form of a (Definition RREF [63]).

If your matrix has only integer entries (as is the case with many examples and exercises in this book), then row operations might introduce rational numbers ("fractions"). So when you enter your matrix, you need to tell SAGE that rational numbers are allowable in its calculations. This is the advice in Computation R.SAGE [1583] to use the ring QQ. As an illustration create

a = matrix(QQ, [[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]])

and issue the command a.echelon_form(). The result is

$$\begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

However, if we adjust the entry by neglecting to specify $\,QQ$, then SAGE assumes that we only want to work with integers, since every entry of the matrix is an integer. So as an experiment, enter

b = matrix([[1, 2, 3, 4], [5, 6, 7, 8], [9, 10, 11, 12]])

and issue the command b.echelon_form(). The result is

1	2	3	4
0	4	8	12
1 0 0	0	0	$\begin{bmatrix} 4 \\ 12 \\ 0 \end{bmatrix}$

You can now clearly see Sage's reluctance to multiply row 2 by $\frac{1}{4}$.

The ring QQ will of course suffice if your matrix has rational numbers for entries. Decimal entries are another place to be careful. If an entry of your matrix is the real number 2.17, you are free to enter it as the rational number $\frac{217}{100}$ and keep the ring QQ in the specification of your matrix. If you want to consider your entries as real numbers, then you might as well just specify your ring as the complex numbers CC. This advice also applies if you have complex numbers as entries.

If you allow SAGE to work with real or complex numbers, then the problem of round-off error becomes relevant. Computer arithmetic with real numbers is, of necessity, subject to minor inaccuracies and errors. This becomes problematic when row-reducing a matrix. If a zero entry is computed instead as an extremely small number, such as 1.287×10^{-18} , then an incorrect sequence of row operations will follow (with further incorrect results). So if you use CC be on the lookout for these kinds of potential pitfalls.

So, in summary, remember to always specify the ring you will be using for your matrices, and most matrices can be handled with a choice of QQ or CC.

When you need to do significant scientific computing with SAGE, there are extra facilities that will help you work with these subtleties.

Finally, you can also use a command of the form a.echelonize() to replace a with its reduced row-echelon form.

Computation Note LS.SAGE Linear Solve

SAGE can solve a variety of systems of equations with the solve() command, even when the equations are not linear (see Exercise SSLE.M70 [41]). But we can afford to specialize here to just linear systems. First, you must specify your variables in advance, so for example, var('x1,x2,x3') might precede a system with three equations. Equations are then written just as you might expect, except that equality is written as ==, since computer programs have traditionally reserved = to assign values to variables. And remember to use a * to indicate that a coefficient multiplies a variable.

The example below illustrates the use of the command and the possibilities for results. Each system would be preceded by establishing the variables with the command var('x,y'). In the case of an infinite solution set, free variables are denoted as rx where x is an integer that increases throughout a session. The style of this description of a solution set

is reminiscent of the style we used in Chapter SLE [2] before we were accustomed to using linear combinations of vectors (Theorem VFSLS [244]).

System	Solution Set	Result
solve([2*x+y==5, 3*x+2*y==15], x, y)	Unique	[[x == -5, y == 15]]
solve([2*x+y==5, 6*x+3*y==15], x, y)	Infinite	[[x == (5 - r1)/2, y == r1]]
solve([2*x+y==5, 6*x+3*y==10], x, y)	Empty	ValueError: Unable to solve

Notice how the output contains equations written a format that might be suitable as input for further use within *SAGE*.

Computation Note VLC.SAGE Vector Linear Combinations

Contributed by Robert Beezer

Vectors in *SAGE* are constructed from lists, and are displayed horizontally. For example, the vector

$$\mathbf{v} = \begin{bmatrix} 1\\ 2\\ 3\\ 4 \end{bmatrix}$$

would be entered and named via the command

v = vector(QQ, [1, 2, 3, 4])

See the notes about rings (Computation R.SAGE [1583]) and matrix entry (Computation ME.SAGE [1584]) for reminders about specifying the relevant ring.

Vector addition and scalar multiplication are then very natural. If $\,u\,$ and $\,v\,$ are two vectors of the same size, then

$$2*u + (-3) * v$$

will compute the correct vector. The result can be assigned to a variable (which will then contain a vector), or be printed. If printed, it will be written horizontally with parentheses for grouping. If u and v have different sizes, then *SAGE* will complain about "unsupported operand(s)."

Computation Note MI.SAGE Matrix Inverse

Contributed by Steve Canfield

If a is a matrix defined in SAGE, then a.inverse() will return the inverse of a, should it exist. In the case where a does not have an inverse SAGE will tell you the matrix must be nonsingular (see Theorem NI [542]).

Computation Note TM.SAGE Transpose of a Matrix

Suppose a is the name of a matrix stored in *SAGE*. Then a.transpose() will return the transpose of a .

Computation Note E.SAGE Eigenspaces

Contributed by Steve Canfield

SAGE can compute eigenspaces and eigenvalues for you. If you have a matrix named **a** and you type

```
a.eigenspaces()
```

you will get a listing of the eigenvalues and the eigenspace for each. Let's do an example. Your output may be formatted slightly different from what we have here.

```
\begin{split} &\texttt{m} = \texttt{matrix}(\texttt{QQ}, [[-13, -8, -4], [12, 7, 4], [24, 16, 7]]) \\ &\texttt{m.eigenspaces}() \\ & [(3, [(1, 2/3, 1/3)]), (-1, [(1, 0, 1/2), (0, 1, -1/2)])] \end{split}
```

Whew, that looks like a mess. At the top level, eigenspaces() returns a dictionary whose keys are the eigenvalues. So in this case we have eigenvalues 3 and -1. Each eigenvalue has an array after it that forms the basis of the eigenspace. In our example, there is 1 vector for $\lambda = 3$ and 2 vectors for $\lambda = -1$. Finally, the vectors SAGE spits out may not be the nicest ones to work with. In particular, we might want to scale the vectors to get rid of fractions.

Appendix P Preliminaries

This appendix contains important ideas about complex numbers, sets, and the logic and techniques of forming proofs. It is not meant to be read straight through, but you should head here when you need to review these ideas.

We choose to expand the set of scalars from the real numbers, \mathbb{R} , to the set of complex numbers, \mathbb{C} . So basic operations with complex numbers (like addition and division) will be necessary. This can be safely postponed until your arrival in Section O [395], and a refresher before Chapter E [944] would be a good idea as well.

Sets are extremely important in all of mathematics, but maybe you have not had much exposure to the basic operations. Check out Section SET [1602]. The text will send you here frequently as well. Visit often.

This book is as much about *doing* mathematics as it is about linear algebra. The "Proof Techniques" are vignettes about logic, types of theorems, structure of proofs, or just plain old-fashioned advice about how to *do* advanced mathematics. The text will frequently point to one of these techniques in advance of their first use, and for specific instructions there will be additional references. If you find constructing proofs difficult (we all did once), then head back here and browse through the advice for second or third readings.

Section CNO Complex Number Operations

In this section we review of the basics of working with complex numbers.

Subsection CNA Arithmetic with complex numbers

A complex number is a linear combination of 1 and $i = \sqrt{-1}$, typically written in the form a + bi. Complex numbers can be added, subtracted, multiplied and divided, just like we are

used to doing with real numbers, including the restriction on division by zero. We will not define these operations carefully, but instead illustrate with examples.

Example ACN Arithmetic of complex numbers

$$(2+5i) + (6-4i) = (2+6) + (5+(-4))i = 8+i$$

$$(2+5i) - (6-4i) = (2-6) + (5-(-4))i = -4+9i$$

$$(2+5i)(6-4i) = (2)(6) + (5i)(6) + (2)(-4i) + (5i)(-4i) = 12+30i - 8i - 20i^{2}$$

$$= 12+22i - 20(-1) = 32+22i$$

Division takes just a bit more care. We multiply the denominator by a complex number chosen to produce a real number and then we can produce a complex number as a result.

$$\frac{2+5i}{6-4i} = \frac{2+5i}{6-4i} \frac{6+4i}{6+4i} = \frac{-8+38i}{52} = -\frac{8}{52} + \frac{38}{52}i = -\frac{2}{13} + \frac{19}{26}i$$

In this example, we used 6+4i to convert the denominator in the fraction to a real number. This number is known as the conjugate, which we define in the next section. We will often exploit the basic properties of complex number addition, subtraction, multiplication and division, so we will carefully define the two basic operations, together with a definition of equality, and then collect nine basic properties in a theorem.

 \boxtimes

Definition CNE **Complex Number Equality**

The complex numbers $\alpha = a + bi$ and $\beta = c + di$ are equal, denoted $\alpha = \beta$, if a = c and b = d. \wedge

(This definition contains Notation CNE.)

Definition CNA

Complex Number Addition

The sum of the complex numbers $\alpha = a + bi$ and $\beta = c + di$, denoted $\alpha + \beta$, is (a+c) + (b+d)i.

(This definition contains Notation CNA.)

Definition CNM **Complex Number Multiplication**

The **product** of the complex numbers $\alpha = a + bi$ and $\beta = c + di$, denoted $\alpha\beta$, is (ac - bd) + bi(ad + bc)i. (This definition contains Notation CNM.) \triangle

Theorem PCNA

Properties of Complex Number Arithmetic

The operations of addition and multiplication of complex numbers have the following properties.

• ACCN Additive Closure, Complex Numbers

If $\alpha, \beta \in \mathbb{C}$, then $\alpha + \beta \in \mathbb{C}$.

 \wedge

- MCCN Multiplicative Closure, Complex Numbers If $\alpha, \beta \in \mathbb{C}$, then $\alpha\beta \in \mathbb{C}$.
- CACN Commutativity of Addition, Complex Numbers For any $\alpha, \beta \in \mathbb{C}, \alpha + \beta = \beta + \alpha$.
- CMCN Commutativity of Multiplication, Complex Numbers For any $\alpha, \beta \in \mathbb{C}, \alpha\beta = \beta\alpha$.
- AACN Additive Associativity, Complex Numbers For any α , β , $\gamma \in \mathbb{C}$, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$.
- MACN Multiplicative Associativity, Complex Numbers For any α , β , $\gamma \in \mathbb{C}$, $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.
- DCN Distributivity, Complex Numbers For any $\alpha, \beta, \gamma \in \mathbb{C}, \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.
- ZCN Zero, Complex Numbers There is a complex number 0 = 0 + 0i so that for any $\alpha \in \mathbb{C}$, $0 + \alpha = \alpha$.
- OCN One, Complex Numbers There is a complex number 1 = 1 + 0i so that for any $\alpha \in \mathbb{C}$, $1\alpha = \alpha$.
- AICN Additive Inverse, Complex Numbers For every $\alpha \in \mathbb{C}$ there exists $-\alpha \in \mathbb{C}$ so that $\alpha + (-\alpha) = 0$.

• MICN Multiplicative Inverse, Complex Numbers For every $\alpha \in \mathbb{C}$, $\alpha \neq 0$ there exists $\frac{1}{\alpha} \in \mathbb{C}$ so that $\alpha(\frac{1}{\alpha}) = 1$.

Proof We could derive each of these properties of complex numbers with a proof that builds on the identical properties of the real numbers. The only proof that might be at all interesting would be to show Property MICN [1598] since we would need to trot out a conjugate. For this property, and especially for the others, we might be tempted to construct proofs of the identical properties for the reals. This would take us way too far afield, so we will draw a line in the sand right here and just agree that these nine fundamental behaviors are true. OK?

Mostly we have stated these nine properties carefully so that we can make reference to them later in other proofs. So we will be linking back here often.

Subsection CCN Conjugates of Complex Numbers

Definition CCN Conjugate of a Complex Number The conjugate of the complex number $\alpha = a + bi \in \mathbb{C}$ is the complex number $\overline{\alpha} = a - bi$.

 \square

(This definition contains Notation CCN.)

Example CSCN Conjugate of some complex numbers

$$\overline{2+3i} = 2-3i$$
 $\overline{5-4i} = 5+4i$ $\overline{-3+0i} = -3+0i$ $\overline{0+0i} = 0+0i$

Notice how the conjugate of a real number leaves the number unchanged. The conjugate enjoys some basic properties that are useful when we work with linear expressions involving addition and multiplication.

Theorem CCRA Complex Conjugation Respects Addition Suppose that α and β are complex numbers. Then $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$.

Proof Let $\alpha = a + bi$ and $\beta = r + si$. Then

$$\overline{\alpha + \beta} = \overline{(a+r) + (b+s)i} = (a+r) - (b+s)i = (a-bi) + (r-si) = \overline{\alpha} + \overline{\beta}$$

Theorem CCRM Complex Conjugation Respects Multiplication \land

 \boxtimes

 \square

Suppose that α and β are complex numbers. Then $\overline{\alpha\beta} = \overline{\alpha}\overline{\beta}$.

Proof Let $\alpha = a + bi$ and $\beta = r + si$. Then

$$\overline{\alpha\beta} = \overline{(ar-bs) + (as+br)i} = (ar-bs) - (as+br)i$$
$$= (ar - (-b)(-s)) + (a(-s) + (-b)r)i = (a-bi)(r-si) = \overline{\alpha}\overline{\beta}$$

Theorem CCT Complex Conjugation Twice

Suppose that α is a complex number. Then $\overline{\overline{\alpha}} = \alpha$.

Proof Let $\alpha = a + bi$. Then

$$\overline{\overline{\alpha}} = \overline{a - bi} = a - (-bi) = a + bi = \alpha$$

Subsection MCN Modulus of a Complex Number

We define one more operation with complex numbers that may be new to you.

Definition MCN Modulus of a Complex Number

The **modulus** of the complex number $\alpha = a + bi \in \mathbb{C}$, is the nonnegative real number

$$|\alpha| = \sqrt{\overline{\alpha}\alpha} = \sqrt{a^2 + b^2}.$$

Example	MSCN		
Modulus	of some	complex	numbers

$$|2+3i| = \sqrt{13}$$
 $|5-4i| = \sqrt{41}$ $|-3+0i| = 3$ $|0+0i| = 0$

 \wedge

The modulus can be interpreted as a version of the absolute value for complex numbers, as is suggested by the notation employed. You can see this in how |-3| = |-3 + 0i| = 3. Notice too how the modulus of the complex zero, 0 + 0i, has value 0.

Section SET Sets

Definition SET

\mathbf{Set}

A set is an unordered collection of objects. If S is a set and x is an object that is in the set S, we write $x \in S$. If x is not in S, then we write $x \notin S$. We refer to the objects in a set as its elements.

(This definition contains Notation SETM.)

Hard to get much more basic than that. Notice that the objects in a set can be *anything*, and there is no notion of order among the elements of the set. A set can be finite as well as infinite. A set can contain other sets as its objects. At a primitive level, a set is just a way to break up some class of objects into two groupings: those objects in the set, and those objects not in the set.

Example SETM Set membership

From the set of all possible symbols, construct the following set of three symbols,

$$S = \{\blacksquare, \blacklozenge, \bigstar\}$$

Δ

Then the statement $\blacksquare \in S$ is true, while the statement $\blacktriangle \in S$ is false. However, then the statement $\blacktriangle \notin S$ is true.

A portion of a set is known as a subset. Notice how the following definition uses an implication (if whenever...then...). Note too how the definition of a subset relies on the definition of a set through the idea of set membership.

Definition SSET

Subset

If S and T are two sets, then S is a subset of T, written $S \subseteq T$ if whenever $x \in S$ then $x \in T$.

(This definition contains Notation SSET.)

If we want to disallow the possibility that S is the same as T, we use the notation $S \subset T$ and we say that S is a **proper subset** of T. We'll do an example, but first we'll define a special set.

Definition ES Empty Set

The empty set is the set with no elements. It is denoted by \emptyset . (This definition contains Notation ES.)

Example SSET

Subset

If
$$S = \{\blacksquare, \blacklozenge, \bigstar\}, T = \{\bigstar, \blacklozenge\}, R = \{\blacktriangle, \bigstar\}$$
, then
 $T \subseteq S$ $R \not\subseteq T$ $\emptyset \subseteq S$

 \wedge

 \triangle

 $T \subset S \qquad \qquad S \subseteq S \qquad \qquad S \not \subset S$

What does it mean for two sets to be equal? They must be the same. Well, that explanation is not really too helpful, is it? How about: If $A \subseteq B$ and $B \subseteq A$, then A equals B. This gives us something to work with, if A is a subset of B, and vice versa, then they must really be the same set. We will now make the symbol "=" do double-duty and extend its use to statements like A = B, where A and B are sets. Here's the definition, which we will reference often.

Definition SE

Set Equality

Two sets, S and T, are equal, if $S \subseteq T$ and $T \subseteq S$. In this case, we write S = T. (This definition contains Notation SE.)

Sets are typically written inside of braces, as { }, as we have seen above. However, when sets have more than a few elements, a description will typically have two components. The first is a description of the general type of objects contained in a set, while the second is some sort of restriction on the properties the objects have. Every object in the set must be of the type described in the first part and it must satisfy the restrictions in the second part. Conversely, any object of the proper type for the first part, that also meets the conditions of the second part, will be in the set. These two parts are set off from each other somehow, often with a vertical bar (|) or a colon (:).

I like to think of sets as clubs. The first part is some description of the type of people who *might* belong to the club, the basic objects. For example, a bicycle club would describe

 \triangle

Version 2.99

its members as being people who like to ride bicycles. The second part is like a membership committee, it restricts the people who are allowed in the club. Continuing with our bicycle club analogy, we might decide to limit ourselves to "serious" riders and only have members who can document having ridden 100 kilometers or more in a single day at least one time.

The restrictions on membership can migrate around some between the first and second part, and there may be several ways to describe the same set of objects. Here's a more mathematical example, employing the set of all integers, \mathbb{Z} , to describe the set of even integers.

 $E = \{ x \in \mathbb{Z} | x \text{ is an even number} \}$ $= \{ x \in \mathbb{Z} | 2 \text{ divides } x \text{ evenly} \}$ $= \{ 2k | k \in \mathbb{Z} \}$

Notice how this set tells us that its objects are integer numbers (not, say, matrices or functions, for example) and just those that are even. So we can write that $10 \in E$, while $17 \notin E$ once we check the membership criteria. We also recognize the question

$$\begin{bmatrix} 1 & -3 & 5 \\ 2 & 0 & 3 \end{bmatrix} \in E?$$

as being simply ridiculous.

Subsection SC Set Cardinality

On occasion, we will be interested in the number of elements in a finite set. Here's the definition and the associated notation.

Definition C Cardinality

Suppose S is a finite set. Then the number of elements in S is called the **cardinality** or size of S, and is denoted |S|.

(This definition contains Notation C.)

 \triangle

 \boxtimes

Example CS Cardinality and Size If $S = \{ \blacklozenge, \bigstar, \blacksquare \}$, then |S| = 3.

Subsection SO Set Operations

In this subsection we define and illustrate the three most common basic ways to manipulate sets to create other sets. Since much of linear algebra is about sets, we will use these often.

Definition SU

Set Union

Suppose S and T are sets. Then the **union** of S and T, denoted $S \cup T$, is the set whose elements are those that are elements of S or of T, or both. More formally,

 $x \in S \cup T$ if and only if $x \in S$ or $x \in T$

(This definition contains Notation SU.)

Notice that the use of the word "or" in this definition is meant to be non-exclusive. That is, it allows for x to be an element of both S and T and still qualify for membership in $S \cup T$.

Example SU Set union If $S = \{ \blacklozenge, \bigstar, \blacksquare \}$ and $T = \{ \diamondsuit, \bigstar, \blacktriangle \}$ then $S \cup T = \{ \diamondsuit, \bigstar, \blacksquare, \blacktriangle \}$.

Definition SI

 \triangle

Set Intersection

Suppose S and T are sets. Then the **intersection** of S and T, denoted $S \cap T$, is the set whose elements are only those that are elements of S and of T. More formally,

 $x \in S \cap T$ if and only if $x \in S$ and $x \in T$

(This definition contains Notation SI.)

Example SI

Set intersection

If $S = \{ \blacklozenge, \bigstar, \blacksquare \}$ and $T = \{ \diamondsuit, \bigstar, \blacktriangle \}$ then $S \cap T = \{ \diamondsuit, \bigstar \}$.

The union and intersection of sets are operations that begin with two sets and produce a third, new, set. Our final operation is the set complement, which we usually think of as an operation that takes a single set and creates a second, new, set. However, if you study the definition carefully, you will see that it needs to be computed *relative* to some "universal" set.

Definition SC

Set Complement

Suppose S is a set that is a subset of a universal set U. Then the **complement** of S, denoted \overline{S} , is the set whose elements are those that are elements of U and not elements of S. More formally,

$$x \in \overline{S}$$
 if and only if $x \in U$ and $x \notin S$

 \triangle

 \boxtimes

(This definition contains Notation SC.)

Notice that there is nothing at all special about the universal set. This is simply a term that suggests that U contains all of the possible objects we are considering. Often this set will be clear from the context, and we won't think much about it, nor reference it in our notation. In other cases (rarely in our work in this course) the exact nature of the universal set must be made explicit, and reference to it will possibly be carried through in our choice of notation.

Example SC

Set complement

If
$$U = \{ \blacklozenge, \bigstar, \blacksquare, \blacktriangle \}$$
 and $S = \{ \diamondsuit, \bigstar, \blacksquare \}$ then $\overline{S} = \{ \blacktriangle \}$.

There are many more natural operations that can be performed on sets, such as an exclusive-or and the symmetric difference. Many of these can be defined in terms of the union, intersection and complement. We will not have much need of them in this course, and so we will not give precise descriptions here in this preliminary section.

There is also an interesting variety of basic results that describe the interplay of these operations with each other. We mention just two as an example, these are known as De-Morgan's Laws.

$$\overline{(S \cup T)} = \overline{S} \cap \overline{T}$$
$$\overline{(S \cap T)} = \overline{S} \cup \overline{T}$$

 \triangle

 \boxtimes

Besides having an appealing symmetry, we mention these two facts, since constructing the proofs of each is a useful exercise that will require a solid understanding of all but one of the definitions presented in this section. Give it a try.

Section PT Proof Techniques

In this section we collect many short essays designed to help you understand how to read, understand and construct proofs. Some are very factual, while others consist of advice. They appear in the order that they are first needed (or advisable) in the text, and are meant to be self-contained. So you should not think of reading through this section in one sitting as you begin this course. But be sure to head back here for a first reading whenever the text suggests it. Also think about returning to browse at various points during the course, and especially as you struggle with becoming an accomplished mathematician who is comfortable with the difficult process of designing new proofs.

Proof Technique D Definitions

A definition is a made-up term, used as a kind of shortcut for some typically more complicated idea. For example, we say a whole number is **even** as a shortcut for saying that when we divide the number by two we get a remainder of zero. With a precise definition, we can

answer certain questions unambiguously. For example, did you ever wonder if zero was an even number? Now the answer should be clear since we have a precise definition of what we mean by the term even.

A single term might have several possible definitions. For example, we could say that the whole number n is even if there is another whole number k such that n = 2k. We say this is an equivalent definition since it categorizes even numbers the same way our first definition does.

Definitions are like two-way streets — we can use a definition to replace something rather complicated by its definition (if it fits) *and* we can replace a definition by its more complicated description. A definition is usually written as some form of an implication, such as "If something-nice-happens, then **blatzo**." However, this also means that "If blatzo, then something-nice-happens," even though this may not be formally stated. This is what we mean when we say a definition is a two-way street — it is really two implications, going in opposite "directions."

Anybody (including you) can make up a definition, so long as it is unambiguous, but the real test of a definition's utility is whether or not it is useful for describing interesting or frequent situations.

We will talk about theorems later (and especially equivalences). For now, be sure not to confuse the notion of a definition with that of a theorem.

In this book, we will display every new definition carefully set-off from the text, and the term being defined will be written thus: **definition**. Additionally, there is a full list of all the definitions, in order of their appearance located at the front of the book (Definitions [ix]). Finally, the acronym for each definition can be found in the index (Index [??]). Definitions

are critical to doing mathematics and proving theorems, so we've given you lots of ways to locate a definition should you forget its...uh, well, ... definition.

Can you formulate a precise definition for what it means for a number to be **odd**? (Don't just say it is the opposite of even. Act as if you don't have a definition for even yet.) Can you formulate your definition a second, equivalent, way? Can you employ your definition to test an odd and an even number for "odd-ness"?

Proof Technique T Theorems

Higher mathematics is about understanding theorems. Reading them, understanding them, applying them, proving them. Every theorem is a shortcut — we prove something in general, and then whenever we find a specific instance covered by the theorem we can immediately say that we know something else about the situation by applying the theorem. In many cases, this new information can be gained with much less effort than if we did not know the theorem.

The first step in understanding a theorem is to realize that the statement of every theorem can be rewritten using statements of the form "If something-happens, then somethingelse-happens." The "something-happens" part is the **hypothesis** and the "something-elsehappens" is the **conclusion**. To understand a theorem, it helps to rewrite its statement using this construction. To apply a theorem, we verify that "something-happens" in a particular instance and immediately conclude that "something-else-happens." To prove a theorem, we must argue based on the assumption that the hypothesis is true, and arrive through the process of logic that the conclusion must then also be true.

Proof Technique L Language

Like any science, the language of math must be understood before further study can continue.

Erin Wilson, Student September, 2004

Mathematics is a language. It is a way to express complicated ideas clearly, precisely, and unambiguously. Because of this, it can be difficult to read. Read slowly, and have pencil and paper at hand. It will usually be necessary to read something several times. While reading can be difficult, it is even harder to speak mathematics, and so that is the topic of this technique.

"Natural" language, in the present case English, is fraught with ambiguity. Consider the possible meanings of the sentence: The fish is ready to eat. One fish, or two fish? Are the fish hungry, or will the fish be eaten? (See Exercise SSLE.M10 [39], Exercise SSLE.M11 [40],

Exercise SSLE.M12 [40], Exercise SSLE.M13 [40].) In your daily interactions with others, give some thought to how many mis-understandings arise from the ambiguity of pronouns, modifiers and objects.

I am going to suggest a simple modification to the way you use language that will make it much, much easier to become proficient at speaking mathematics and eventually it will become second nature. Think of it as a training aid or practice drill you might use when learning to become skilled at a sport.

First, eliminate pronouns from your vocabulary when discussing linear algebra, in class or with your colleagues. Do not use: it, that, those, their or similar sources of confusion. This is the single easiest step you can take to make your oral expression of mathematics clearer to others, and in turn, it will greatly help your own understanding.

Now rid yourself of the word "thing" (or variants like "something"). When you are tempted to use this word realize that there is some object you want to discuss, and we likely have a definition for that object (see the discussion at Technique D [1611]). Always "think about your objects" and many aspects of the study of mathematics will get easier. Ask yourself: "Am I working with a set, a number, a function, an operation, a differential equation, or what?" Knowing what an object *is* will allow you to narrow down the procedures you may apply to **it**. If you have studied an object-oriented computer programming language, then you will already have experience identifying objects and thinking carefully about what procedures are allowed to be applied to them.

Third, eliminate the verb "works" (as in "the equation works") from your vocabulary. This term is used as a substitute when we are not sure just what we are trying to accomplish. Usually we are trying to say that some object fulfills some condition. The condition might even have a definition associated with it, making it even easier to describe.

Last, speak sloooowly and thoughtfully as you try to get by without all these lazy words. It is hard at first, but you will get better with practice. Especially in class, when the pressure is on and all eyes are on you, don't succumb to the temptation to use these weak words. Slow down, we'd all rather wait for a slow, well-formed question or answer than a fast, sloppy, incomprehensible one.

You will find the improvement in your ability to *speak* clearly about complicated ideas will greatly improve your ability to *think* clearly about complicated ideas. And I believe that you cannot think clearly about complicated ideas if you cannot formulate questions or answers clearly in the correct language. This is as applicable to the study of law, economics or philosophy as it is to the study of science or mathematics.

In this spirit, Dupont Hubert has contributed the following quotation, which is widely used in French mathematics courses (and which might be construed as the contrapositive of Technique CP [1621])

Ce que l'on concoit bien s'enonce clairement, Et les mots pour le dire arrivent aisement.

— Nicolas Boileau, L'art poétique, Chant I, 1674

which translates as

Whatever is well conceived is clearly said, And the words to say it flow with ease. So when you come to class, check your pronouns at the door, along with other weak words. And when studying with friends, you might make a game of catching one another using pronouns, "thing," or "works." I know I'll be calling you on it!

Proof Technique GS Getting Started

"I don't know how to get started!" is often the lament of the novice proof-builder. Here are a few pieces of advice.

- 1. As mentioned in Technique T [1613], rewrite the statement of the theorem in an "ifthen" form. This will simplify identifying the hypothesis and conclusion, which are referenced in the next few items.
- 2. Ask yourself what *kind* of statement you are trying to prove. This is always part of your conclusion. Are you being asked to conclude that two numbers are equal, that a function is differentiable or a set is a subset of another? You cannot bring other techniques to bear if you do not know what *type* of conclusion you have.
- 3. Write down reformulations of your hypotheses. Interpret and translate each definition properly.

- 4. Write your hypothesis at the top of a sheet of paper and your conclusion at the bottom. See if you can formulate a statement that precedes the conclusion and also implies it. Work down from your hypothesis, and up from your conclusion, and see if you can meet in the middle. When you are finished, rewrite the proof nicely, from hypothesis to conclusion, with verifiable implications giving each subsequent statement.
- 5. As you work through your proof, think about what kinds of objects your symbols represent. For example, suppose A is a set and f(x) is a real-valued function. Then the expression A + f might make no sense if we have not defined what it means to "add" a set to a function, so we can stop at that point and adjust accordingly. On the other hand we might understand 2f to be the function whose rule is described by (2f)(x) = 2f(x). "Think about your objects" means to always verify that your objects and operations are compatible.

Proof Technique C Constructive Proofs

Conclusions of proofs come in a variety of types. Often a theorem will simply *assert* that something exists. The best way, but not the only way, to show something exists is to actually build it. Such a proof is called **constructive**. The thing to realize about constructive proofs is that the proof itself will contain a procedure that might be used computationally to construct the desired object. If the procedure is not too cumbersome, then the proof itself is as useful as the statement of the theorem.

Proof Technique E Equivalences

When a theorem uses the phrase "if and only if" (or the abbreviation "iff") it is a shorthand way of saying that two if-then statements are true. So if a theorem says "P if and only if Q," then it is true that "if P, then Q" while it is also true that "if Q, then P." For example, it may be a theorem that "I wear bright yellow knee-high plastic boots if and only if it is raining." This means that I *never* forget to wear my super-duper yellow boots when it is raining *and* I wouldn't be seen in such silly boots *unless* it was raining. You never have one without the other. I've got my boots on and it is raining *or* I don't have my boots on and it is dry.

The upshot for proving such theorems is that it is like a 2-for-1 sale, we get to do *two* proofs. Assume P and conclude Q, then start over and assume Q and conclude P. For this reason, "if and only if" is sometimes abbreviated by \iff , while proofs indicate which of the two implications is being proved by prefacing each with \Rightarrow or \Leftarrow . A carefully written proof will remind the reader which statement is being used as the hypothesis, a quicker version will let the reader deduce it from the direction of the arrow. Tradition dictates we do the "easy" half first, but that's hard for a student to know until you've finished doing both halves! Oh

well, if you rewrite your proofs (a good habit), you can then choose to put the easy half first.

Theorems of this type are called "equivalences" or "characterizations," and they are some of the most pleasing results in mathematics. They say that two objects, or two situations, are really the same. You don't have one without the other, like rain and my yellow boots. The more different P and Q seem to be, the more pleasing it is to discover they are really equivalent. And if P describes a very mysterious solution or involves a tough computation, while Q is transparent or involves easy computations, then we've found a great shortcut for better understanding or faster computation. Remember that every theorem really is a shortcut in some form. You will also discover that if proving $P \Rightarrow Q$ is very easy, then proving $Q \Rightarrow P$ is likely to be proportionately harder. Sometimes the two halves are about equally hard. And in rare cases, you can string together a whole sequence of other equivalences to form the one you're after and you don't even need to do two halves. In this case, the argument of one half is just the argument of the other half, but in reverse.

One last thing about equivalences. If you see a statement of a theorem that says two things are "equivalent," translate it first into an "if and only if" statement.

Proof Technique N Negation

When we construct the contrapositive of a theorem (Technique CP [1621]), we need to negate the two statements in the implication. And when we construct a proof by contra-

diction (Technique CD [1623]), we need to negate the conclusion of the theorem. One way to construct a converse (Technique CV [1622]) is to simultaneously negate the hypothesis and conclusion of an implication (but remember that this is not guaranteed to be a true statement). So we often have the need to negate statements, and in some situations it can be tricky.

If a statement says that a set is empty, then its negation is the statement that the set is nonempty. That's straightforward. Suppose a statement says "something-happens" for all i, or every i, or any i. Then the negation is that "something-doesn't-happen" for at least one value of i. If a statement says that there exists at least one "thing," then the negation is the statement that there is no "thing." If a statement says that a "thing" is unique, then the negation is that there is zero, or more than one, of the "thing."

We are not covering all of the possibilities, but we wish to make the point that logical qualifiers like "there exists" or "for every" must be handled with care when negating statements. Studying the proofs which employ contradiction (as listed in Technique CD [1623]) is a good first step towards understanding the range of possibilities.

Proof Technique CP Contrapositives

The **contrapositive** of an implication $P \Rightarrow Q$ is the implication $not(Q) \Rightarrow not(P)$, where "not" means the logical negation, or opposite. An implication is true if and only if its contrapositive is true. In symbols, $(P \Rightarrow Q) \iff (\operatorname{not}(Q) \Rightarrow \operatorname{not}(P))$ is a theorem. Such statements about logic, that are always true, are known as **tautologies**.

For example, it is a theorem that "if a vehicle is a fire truck, then it has big tires and has a siren." (Yes, I'm sure you can conjure up a counterexample, but play along with me anyway.) The contrapositive is "if a vehicle does not have big tires or does not have a siren, then it is not a fire truck." Notice how the "and" became an "or" when we negated the conclusion of the original theorem.

It will frequently happen that it is easier to construct a proof of the contrapositive than of the original implication. If you are having difficulty formulating a proof of some implication, see if the contrapositive is easier for you. The trick is to construct the negation of complicated statements accurately. More on that later.

Proof Technique CV Converses

The **converse** of the implication $P \Rightarrow Q$ is the implication $Q \Rightarrow P$. There is no guarantee that the truth of these two statements are related. In particular, if an implication has been proven to be a theorem, then do not try to use its converse too, as if it were a theorem. Sometimes the converse is true (and we have an equivalence, see Technique E [1619]). But more likely the converse is false, especially if it wasn't included in the statement of the original theorem.

For example, we have the theorem, "if a vehicle is a fire truck, then it is has big tires and has a siren." The converse is false. The statement that "if a vehicle has big tires and a siren, then it is a fire truck" is false. A police vehicle for use on a sandy public beach would have big tires and a siren, yet is not equipped to fight fires.

We bring this up now, because Theorem CSRN [123] has a tempting converse. Does this theorem say that if r < n, then the system is consistent? Definitely not, as Archetype E [1678] has r = 3 < 4 = n, yet is inconsistent. This example is then said to be a **counterexample** to the converse. Whenever you think a theorem that is an implication might actually be an equivalence, it is good to hunt around for a counterexample that shows the converse to be false (the archetypes, Appendix A [1637], can be a good hunting ground).

Proof Technique CD Contradiction

Another proof technique is known as "proof by contradiction" and it can be a powerful (and satisfying) approach. Simply put, suppose you wish to prove the implication, "If A, then B." As usual, we assume that A is true, but we also make the additional assumption that B is false. If our original implication is true, then these twin assumptions should lead us to a logical inconsistency. In practice we assume the negation of B to be true (see Technique N [1620]). So we argue from the assumptions A and not(B) looking for some obviously false conclusion such as 1 = 6, or a set is simultaneously empty and nonempty, or a matrix is

both nonsingular and singular.

You should be careful about formulating proofs that look like proofs by contradiction, but really aren't. This happens when you assume A and not(B) and proceed to give a "normal" and direct proof that B is true by only using the assumption that A is true. Your last step is to then claim that B is true and you then appeal to the assumption that not(B) is true, thus getting the desired contradiction. Instead, you could have avoided the overhead of a proof by contradiction and just run with the direct proof. This stylistic flaw is known, quite graphically, as "setting up the strawman to knock him down."

Here is a simple example of a proof by contradiction. There are direct proofs that are just about as easy, but this will demonstrate the point, while narrowly avoiding knocking down the straw man.

Theorem: If a and b are odd integers, then their product, ab, is odd.

Proof: To begin a proof by contradiction, assume the hypothesis, that a and b are odd. Also assume the negation of the conclusion, in this case, that ab is even. Then there are integers, j, k, ℓ so that a = 2j + 1, b = 2k + 1, $ab = 2\ell$. Then

$$0 = ab - ab$$

= $(2j + 1)(2k + 1) - (2\ell)$
= $4jk + 2j + 2k - 2\ell + 1$
= $2(2jk + j + k - \ell) + 1$

Notice how we used both our hypothesis and the negation of the conclusion in the second line. Now divide the integer on each end of this string of equalities by 2. On the left we get a remainder of 0, while on the right we see that the remainder will be 1. Both remainders cannot be correct, so this is our desired contradiction. Thus, the conclusion (that ab is odd) is true.

Again, we do not offer this example as the *best* proof of this fact about even and odd numbers, but rather it is a simple illustration of a proof by contradiction. You can find examples of proofs by contradiction in Theorem RREFU [70], Theorem NMUS [180], Theorem NPNT [538], Theorem TTMI [509], Theorem GSP [413], Theorem ELIS [850], Theorem EDYES [858], Theorem EMHE [953], Theorem EDELI [1001], and Theorem DMFE [1044], in addition to several examples and solutions to exercises.

Proof Technique U Uniqueness

A theorem will sometimes claim that some object, having some desirable property, is unique. In other words, there should be only one such object. To prove this, a standard technique is to assume there are two such objects and proceed to analyze the consequences. The end result may be a contradiction (Technique CD [1623]), or the conclusion that the two allegedly different objects really are equal.

Proof Technique ME Multiple Equivalences

A very specialized form of a theorem begins with the statement "The following are equivalent...," which is then followed by a list of statements. Informally, this lead-in sometimes gets abbreviated by "TFAE." This formulation means that any two of the statements on the list can be connected with an "if and only if" to form a theorem. So if the list has n statements then, there are $\frac{n(n-1)}{2}$ possible equivalences that can be constructed (and are claimed to be true).

Suppose a theorem of this form has statements denoted as $A, B, C, \ldots Z$. To prove the entire theorem, we can prove $A \Rightarrow B, B \Rightarrow C, C \Rightarrow D, \ldots, Y \Rightarrow Z$ and finally, $Z \Rightarrow A$. This circular chain of n equivalences would allow us, logically, if not practically, to form any one of the $\frac{n(n-1)}{2}$ possible equivalences by chasing the equivalences around the circle as far as required.

Proof Technique PI Proving Identities

Many theorems have conclusions that say two objects are equal. Perhaps one object is hard to compute or understand, while the other is easy to compute or understand. This would make for a pleasing theorem. Whether the result is pleasing or not, we take the same approach to formulate a proof. Sometimes we need to employ specialized notions of equality, such as Definition SE [1604] or Definition CVE [199], but in other cases we can string together a list of equalities.

The wrong way to prove an identity is to begin by writing it down and then beating on it until it reduces to an obvious identity. The first flaw is that you would be writing down the statement you wish to prove, as if you already believed it to be true. But more dangerous is the possibility that some of your maneuvers are not reversible. Here's an example. Let's prove that 3 = -3.

3 = -3	(This is a bad start)
$3^2 = (-3)^2$	Square both sides
9 = 9	
0 = 0	Subtract 9 from both sides

So because 0 = 0 is a true statement, does it follow that 3 = -3 is a true statement? Nope.

Of course, we didn't really expect a legitimate proof of 3 = -3, but this attempt should illustrate the dangers of this (incorrect) approach.

What you have just seen in the proof of Theorem VSPCV [205], and what you will see consistently throughout this text, is proofs of the following form. To prove that A = D we write

A = B	Theorem, Definition or Hypothesis justifying $A = B$
= C	Theorem, Definition or Hypothesis justifying $B = C$
= D	Theorem, Definition or Hypothesis justifying $C = D$

In your scratch work exploring possible approaches to proving a theorem you may massage a variety of expressions, sometimes making connections to various bits and pieces, while some parts get abandoned. Once you see a line of attack, rewrite your proof carefully mimicking this style.

Proof Technique DC Decompositions

Much of your mathematical upbringing, especially once you began a study of algebra, revolved around simplifying expressions — combining like terms, obtaining common denominators so as to add fractions, factoring in order to solve polynomial equations. However, as often as not, we will do the opposite. Many theorems and techniques will revolve around taking some object and "decomposing" it into some combination of other objects, ostensibly in a more complicated fashion. When we say something can "be written as" something else, we mean that the one object can be decomposed into some combination of other objects. This may seem unnatural at first, but results of this type will give us insight into the structure of the original object by exposing its inner workings. An appropriate analogy might be stripping the wallboards away from the interior of a building to expose the structural members supporting the whole building.

Perhaps you have studied integral calculus, or a pre-calculus course, where you learned about partial fractions. This is a technique where a fraction of two polynomials is *decomposed* (written as, expressed as) a sum of simpler fractions. The purpose in calculus is to make finding an antiderivative simpler. For example, you can verify the truth of the expression

$$\frac{12x^5 + 2x^4 - 20x^3 + 66x^2 - 294x + 308}{x^6 + x^5 - 3x^4 + 21x^3 - 52x^2 + 20x - 48} = \frac{5x + 2}{x^2 - x + 6} + \frac{3x - 7}{x^2 + 1} + \frac{3}{x + 4} + \frac{1}{x - 2}$$

In an early course in algebra, you might be expected to combine the four terms on the right over a common denominator to create the "simpler" expression on the left. Going the other way, the partial fraction technique would allow you to systematically *decompose* the fraction of polynomials on the left into the sum of the four (arguably) simpler fractions of polynomials on the right.

This is a major shift in thinking, so come back here often, especially when we say "can be written as", or "can be expressed as," or "can be decomposed as."

Proof Technique I Induction

"Induction" or "mathematical induction" is a framework for proving statements that are indexed by integers. In other words, suppose you have a statement to prove that is really multiple statements, one for n = 1, another for n = 2, a third for n = 3, and so on. If there is enough similarity between the statements, then you can use a script (the framework) to prove them all at once.

For example, consider the theorem

Theorem
$$1 + 2 + 3 + \dots + n = \frac{n(n+1)}{2}$$
 for $n \ge 1$.

This is shorthand for the many statements $1 = \frac{1(1+1)}{2}$, $1+2 = \frac{2(2+1)}{2}$, $1+2+3 = \frac{3(3+1)}{2}$, $1+2+3+4 = \frac{4(4+1)}{2}$, and so on. Forever. You can do the calculations in each of these statements and verify that all four are true. We might not be surprised to learn that the fifth statement is true as well (go ahead and check). However, do we think the theorem is true for n = 872? Or n = 1, 234, 529?

To see that these questions are not so ridiculous, consider the following example from Rotman's Journey into Mathematics. The statement " n^2-n+41 is prime" is true for integers $1 \le n \le 40$ (check a few). However, when we check n = 41 we find $41^2 - 41 + 41 = 41^2$, which is not prime.

So how do we prove infinitely many statements all at once? More formally, lets denote our statements as P(n). Then, if we can prove the two assertions

1. P(1) is true.

2. If P(k) is true, then P(k+1) is true.

then it follows that P(n) is true for all $n \ge 1$. To understand this, I liken the process to climbing an infinitely long ladder with equally spaced rungs. Confronted with such a ladder, suppose I tell you that you are able to step up onto the first rung, and if you are on any particular rung, then you are capable of stepping up to the next rung. It follows that you can climb the ladder as far up as you wish. The first formal assertion above is akin to stepping onto the first rung, and the second formal assertion is akin to assuming that if you are on any one rung then you can always reach the next rung.

In practice, establishing that P(1) is true is called the "base case" and in most cases is straightforward. Establishing that $P(k) \Rightarrow P(k+1)$ is referred to as the "induction step," or in this book (and elsewhere) we will typically refer to the assumption of P(k) as the "induction hypothesis." This is perhaps the most mysterious part of a proof by induction, since it looks like you are assuming (P(k)) what you are trying to prove (P(n)). Sometimes it is even worse, since as you get more comfortable with induction, we often don't bother to use a different letter (k) for the index (n) in the induction step. Notice that the second formal assertion never says that P(k) is true, it simply says that if P(k) were true, what might logically follow. We can establish statements like "If I lived on the moon, then I could pole-vault over a bar 12 meters high." This may be a true statement, but it does not say we live on the moon, and indeed we may never live there. Enough generalities. Let's work an example and prove the theorem above about sums of integers. Formally, our statement is P(n): $1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}$.

Proof: Base Case. P(1) is the statement $1 = \frac{1(1+1)}{2}$, which we see simplifies to the true statement 1 = 1.

Induction Step: We will assume P(k) is true, and will try to prove P(k+1). Given what we want to accomplish, it is natural to begin by examining the sum of the first k+1 integers.

$$1 + 2 + 3 + \dots + (k + 1) = (1 + 2 + 3 + \dots + k) + (k + 1)$$

= $\frac{k(k + 1)}{2} + (k + 1)$ Induction Hypothesis
= $\frac{k^2 + k}{2} + \frac{2k + 2}{2}$
= $\frac{k^2 + 3k + 2}{2}$
= $\frac{(k + 1)(k + 2)}{2}$
= $\frac{(k + 1)((k + 1) + 1)}{2}$

We then recognize the two ends of this chain of equalities as P(k+1). So, by mathematical induction, the theorem is true for all n.

How do you recognize when to use induction? The first clue is a statement that is

really many statements, one for each integer. The second clue would be that you begin a more standard proof and you find yourself using words like "and so on" (as above!) or lots of ellipses (dots) to establish patterns that you are convinced continue on and on forever. However, there are many minor instances where induction might be warranted but we don't bother.

Induction is important enough, and used often enough, that it appears in various variations. The base case sometimes begins with n = 0, or perhaps an integer greater than n. Some formulate the induction step as $P(k-1) \Rightarrow P(k)$. There is also a "strong form" of induction where we assume all of P(1), P(2), P(3), ... P(k) as a hypothesis for showing the conclusion P(k+1).

You can find examples of induction in the proofs of Theorem GSP [413], Theorem DER [897], Theorem DT [899], Theorem DIM [930], Theorem EOMP [1006], Theorem DCP [1013], and Theorem KPLT [1450].

Proof Technique P Practice

Here is a technique used by many practicing mathematicians when they are teaching themselves new mathematics. As they read a textbook, monograph or research article, they attempt to prove each new theorem themselves, *before* reading the proof. Often the proofs can be very difficult, so it is wise not to spend too much time on each. Maybe limit your losses and try each proof for 10 or 15 minutes. Even if the proof is not found, it is time well-spent. You become more familiar with the definitions involved, and the hypothesis and conclusion of the theorem. When you do work through the proof, it might make more sense, and you will gain added insight about just how to construct a proof.

Proof Technique LC Lemmas and Corollaries

Theorems often go by different titles. Two of the most popular being "lemma" and "corollary." Before we describe the fine distinctions, be aware that lemmas, corollaries, propositions, claims and facts are all just theorems. And every theorem can be rephrased as an "if-then" statement, or perhaps a pair of "if-then" statements expressed as an equivalence (Technique E [1619]).

A lemma is a theorem that is not too interesting in its own right, but is important for proving other theorems. It might be a generalization or abstraction of a key step of several different proofs. For this reason you often hear the phrase "technical lemma" though some might argue that the adjective "technical" is redundant.

A corollary is a theorem that follows very easily from another theorem. For this reason, corollaries frequently do not have proofs. You are expected to easily and quickly see how a previous theorem implies the corollary.

A proposition or fact is really just a codeword for a theorem. A claim might be similar,

but some authors like to use claims within a proof to organize key steps. In a similar manner, some long proofs are organized as a series of lemmas.

In order to not confuse the novice, we have just called all our theorems theorems. It is also an organizational convenience. With only theorems and definitions, the theoretical backbone of the course is laid bare in the two lists of Definitions [ix] and Theorems [x].

Appendix A Archetypes

WordNet (an open-source lexical database) gives the following definition of "archetype": something that serves as a model or a basis for making copies.

Our archetypes are typical examples of systems of equations, matrices and linear transformations. They have been designed to demonstrate the range of possibilities, allowing you to compare and contrast them. Several are of a size and complexity that is usually not presented in a textbook, but should do a better job of being "typical."

We have made frequent reference to many of these throughout the text, such as the frequent comparisons between Archetype A [1642] and Archetype B [1652]. Some we have left for you to investigate, such as Archetype J [1723], which parallels Archetype I [1714].

How should you use the archetypes? First, consult the description of each one as it is

mentioned in the text. See how other facts about the example might illuminate whatever property or construction is being described in the example. Second, each property has a short description that usually includes references to the relevant theorems. Perform the computations and understand the connections to the listed theorems. Third, each property has a small checkbox in front of it. Use the archetypes like a workbook and chart your progress by "checking-off" those properties that you understand.

The next page has a chart that summarizes some (but not all) of the properties described for each archetype. Notice that while there are several types of objects, there are fundamental connections between them. That some lines of the table do double-duty is meant to convey some of these connections. Consult this table when you wish to quickly find an example of a certain phenomenon.

H		цЭ	цЭ		(r)	64						\bigcirc			Ap	penc	lix	А	Arc	hety	\mathbf{pes}	1643	;
K	Σ	Ŋ	Ŋ		Ŋ	0			Х	Х	Х	16	Х										-
F	∞	6	9	Ι	4	ю			Z							ions							
Π	∞	7	4	Ι	3	4			Ζ						-	mat: olut	יורט ארט	arr					
Н	∞	2	Ŋ	Z	2	0			Х						ر	No s		IICAL					
Ċ	∞	2	Ŋ	D	2	0			Х						F	Lran N≡		ddv					
Ĺ	∞	4	4	Ŋ	4	0			Х	Х	Υ	-18	Х			near ions	-Not	1011-					
ГÌ	∞	4	ŝ	Z	2	2			Z					cts		illut solut		– VIII					
D	∞	4	ŝ	Π	2	2			Z					Ч	i F	х, г		, DIG					
U	∞	4	ŝ	Η	က				Х					/De	, . , .	r me	ible	DIUIC					
В	∞	က	က	D	က	0			Х	Х	Х	-2	Х	Archetype Facts		1≡N iteli	ionit	end r					
A	∞	S	ŝ	Ι	2				Ζ	Ζ	Z	0	Z	Arcl		ns, n Infir							
	Type	Vars, Cols, Domain	Eqns, Rows, CoDom	Solution Set	Rank	Nullity	Injective	Surjective	Full Rank	Nonsingular	Invertible	Determinant	Diagonalizable	4		D=Dystem of Equations, M=Matrix, L=Linear Iransformation II=Unique solution I=Infinitely many solutions N=No solutions	V-Vindue Bottanon, 1-minimuter, manifestation and 2000 Volume V-Voc N-No V-Immeetikle klank-Net Amiliaakle	I — I CD, IN — INO, V					

Archetype A

Summary Linear system of three equations, three unknowns. Singular coefficient matrix with dimension 1 null space. Integer eigenvalues and a degenerate eigenspace for coefficient matrix.

A system of linear equations (Definition SLE [18]):

 $x_1 - x_2 + 2x_3 = 1$ $2x_1 + x_2 + x_3 = 8$ $x_1 + x_2 = 5$

Some solutions to the system of linear equations (not necessarily exhaustive):

 $x_1 = 2, \quad x_2 = 3, \quad x_3 = 1$

$$x_1 = 3, \quad x_2 = 2, \quad x_3 = 0$$

Augmented matrix of the linear system of equations (Definition AM [57]): $\begin{bmatrix}
1 & -1 & 2 & 1 \\
2 & 1 & 1 & 8 \\
1 & 1 & 0 & 5
\end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$\left[1 \right]$	0	1	3
0	1	-1	2
0	0	0	0

Analysis of the augmented matrix (Notation RREFA [64]):

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3, 4\}$

U Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the

pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$x_1 - x_2 + 2x_3 = 0$$

$$2x_1 + x_2 + x_3 = 0$$

$$x_1 + x_2 = 0$$

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

 $x_1 = 0, \quad x_2 = 0, \quad x_3 = 0$

$$x_1 = -1, \quad x_2 = 1, \quad x_3 = 1$$

$$x_1 = -5, \quad x_2 = 5, \quad x_3 = 5$$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3, 4\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.

 $\begin{bmatrix} 1 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

Matrix brought to reduced row-echelon form:

 $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$

Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 2 $D = \{1, 2\}$ $F = \{3\}$

 \square Matrix (coefficient matrix) is nonsingular or singular? (Theorem NMRRI [176]) at the same time, examine the size of the set F above.Notice that this property does not apply to matrices that are not square.

Singular.

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem

SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

$$\left\langle \left\{ \begin{bmatrix} -1\\1\\1 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

 $\left\langle \left\{ \begin{bmatrix} 1\\2\\1 \end{bmatrix}, \begin{bmatrix} -1\\1\\1 \end{bmatrix} \right\} \right\rangle$

The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$$L = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}$$
$$\left\langle \left\{ \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

 $\left\langle \left\{ \begin{bmatrix} 1\\0\\-\frac{1}{3} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{2}{3} \end{bmatrix} \right\} \right\rangle$

Bow space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

$\left\langle \left\{ \begin{bmatrix} 1\\0\\1 \end{bmatrix}, \begin{bmatrix} 0\\1\\-1 \end{bmatrix} \right\} \right\rangle$

Inverse matrix, if it exists. The inverse is not defined for matrices that are not square, and if the matrix is square, then the matrix must be nonsingular. (Definition MI [506], Theorem NI [542])

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 3 Rank: 2 Nullity: 1

Determinant of the matrix, which is only defined for square matrices. The matrix is nonsingular if and only if the determinant is nonzero (Theorem SMZD [935]). (Product of all eigenvalues?)

Determinant = 0

Eigenvalues, and bases for eigenspaces. (Definition EEM [945], Definition EM [963])

$$\lambda = 0 \qquad \qquad \mathcal{E}_A(0) = \left\langle \left\{ \begin{bmatrix} -1\\1\\1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 2 \qquad \qquad \mathcal{E}_A(2) = \left\langle \left\{ \begin{bmatrix} 1\\5\\3 \end{bmatrix} \right\} \right\rangle$$

Geometric and algebraic multiplicities. (Definition GME [969]Definition AME [968])

$$\gamma_A(0) = 1 \qquad \qquad \alpha_A(0) = 2 \gamma_A(2) = 1 \qquad \qquad \alpha_A(2) = 1$$

Diagonalizable? (Definition DZM [1037])

No, $\gamma_A(0) \neq \alpha_B(0)$, Theorem DMFE [1044].

Archetype B

Summary System with three equations, three unknowns. Nonsingular coefficient matrix. Distinct integer eigenvalues for coefficient matrix.

A system of linear equations (Definition SLE [18]):

$$-7x_1 - 6x_2 - 12x_3 = -33$$

$$5x_1 + 5x_2 + 7x_3 = 24$$

$$x_1 + 4x_3 = 5$$

Some solutions to the system of linear equations (not necessarily exhaustive):

$$x_1 = -3, \quad x_2 = 5, \quad x_3 = 2$$

____ Augmented matrix of the linear system of equations (Definition AM [57]):

 $\begin{bmatrix} -7 & -6 & -12 & -33 \\ 5 & 5 & 7 & 24 \\ 1 & 0 & 4 & 5 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$\left[1 \right]$	0	0	-3
0	1	0	5
0	0	1	2

Analysis of the augmented matrix (Notation RREFA [64]):

$$r = 3$$
 $D = \{1, 2, 3\}$ $F = \{4\}$

U Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the

pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -3 \\ 5 \\ 2 \end{bmatrix}$$

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

 $-11x_1 + 2x_2 - 14x_3 = 0$ $23x_1 - 6x_2 + 33x_3 = 0$ $14x_1 - 2x_2 + 17x_3 = 0$

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

 $x_1 = 0, \quad x_2 = 0, \quad x_3 = 0$

Form the augmented matrix of the homogenous linear system, and use row operations

to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

$\boxed{1}$	0	0	0
0	1	0	0
0	0	1	0

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 3$$
 $D = \{1, 2, 3\}$ $F = \{4\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.

$$\begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix}$$

Matrix brought to reduced row-echelon form:

 $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 3 $D = \{1, 2, 3\}$ $F = \{\}$

 \square Matrix (coefficient matrix) is nonsingular or singular? (Theorem NMRRI [176]) at the same time, examine the size of the set F above.Notice that this property does not apply to matrices that are not square.

Nonsingular.

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem

SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

 $\langle \{ \} \rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

$$\left\langle \left\{ \begin{bmatrix} -7\\5\\1 \end{bmatrix}, \begin{bmatrix} -6\\5\\0 \end{bmatrix}, \begin{bmatrix} -12\\7\\4 \end{bmatrix} \right\} \right\rangle$$

The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

 $L = \begin{bmatrix} \end{bmatrix}$

$\left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

 $\left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle$

Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

$\left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle$

Inverse matrix, if it exists. The inverse is not defined for matrices that are not square, and if the matrix is square, then the matrix must be nonsingular. (Definition MI [506], Theorem NI [542])

$\left[-10\right]$	-12	-9]
$\frac{\underline{13}}{\underline{2}}$	8	$\frac{11}{\frac{2}{5}}$
$\frac{5}{2}$	3	$\frac{5}{2}$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 3 Rank: 3 Nullity: 0

Determinant of the matrix, which is only defined for square matrices. The matrix is nonsingular if and only if the determinant is nonzero (Theorem SMZD [935]). (Product of all eigenvalues?)

Determinant = -2

Eigenvalues, and bases for eigenspaces. (Definition EEM [945], Definition EM [963])

.

$$\lambda = -1 \qquad \qquad \mathcal{E}_B(-1) = \left\langle \left\{ \begin{bmatrix} -5\\3\\1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 1 \qquad \qquad \mathcal{E}_B(1) = \left\langle \left\{ \begin{bmatrix} -3\\2\\1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 2 \qquad \qquad \mathcal{E}_B(2) = \left\langle \left\{ \begin{bmatrix} -2\\1\\1 \end{bmatrix} \right\} \right\rangle$$

Geometric and algebraic multiplicities. (Definition GME [969]Definition AME [968])

$$\gamma_B\left(-1\right) = 1 \qquad \qquad \alpha_B\left(-1\right) = 1$$

$$\gamma_B (1) = 1 \qquad \qquad \alpha_B (1) = 1 \gamma_B (2) = 1 \qquad \qquad \alpha_B (2) = 1$$

Diagonalizable? (Definition DZM [1037])

Yes, distinct eigenvalues, Theorem DED [1048].

The diagonalization. (Theorem DC [1039])

$$\begin{bmatrix} -1 & -1 & -1 \\ 2 & 3 & 1 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Archetype C

Summary System with three equations, four variables. Consistent. Null space of coefficient matrix has dimension 1.

A system of linear equations (Definition SLE [18]):

$$2x_1 - 3x_2 + x_3 - 6x_4 = -7$$

$$4x_1 + x_2 + 2x_3 + 9x_4 = -7$$

$$3x_1 + x_2 + x_3 + 8x_4 = -8$$

Some solutions to the system of linear equations (not necessarily exhaustive):

 $x_1 = -7, \quad x_2 = -2, \quad x_3 = 7, \quad x_4 = 1$

$$x_1 = -1, \quad x_2 = 7, \quad x_3 = 4, \quad x_4 = -2$$

Augmented matrix of the linear system of equations (Definition AM [57]): $\begin{bmatrix} 2 & -3 & 1 & -6 & -7 \end{bmatrix}$

 $\begin{bmatrix} 2 & -3 & 1 & -6 & -7 \\ 4 & 1 & 2 & 9 & -7 \\ 3 & 1 & 1 & 8 & -8 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

Γ	1	0	0	2	-5
		1	0	3	1
	0	0	1	-1	6

Analysis of the augmented matrix (Notation RREFA [64]):

$$r = 3$$
 $D = \{1, 2, 3\}$ $F = \{4, 5\}$

 \Box Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

$$\begin{bmatrix} x_1\\x_2\\x_3\\x_4 \end{bmatrix} = \begin{bmatrix} -5\\1\\6\\0 \end{bmatrix} + x_4 \begin{bmatrix} -2\\-3\\1\\1 \end{bmatrix}$$

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

 $2x_1 - 3x_2 + x_3 - 6x_4 = 0$ $4x_1 + x_2 + 2x_3 + 9x_4 = 0$ $3x_1 + x_2 + x_3 + 8x_4 = 0$

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

$$x_1 = 0, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$$

Version 2.99

$$x_1 = -2, \quad x_2 = -3, \quad x_3 = 1, \quad x_4 = 1$$

 $x_1 = -4, \quad x_2 = -6, \quad x_3 = 2, \quad x_4 = 2$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

1	0	0	2	0
0	1	0	3	0
0	0	1	-1	0

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 3$$
 $D = \{1, 2, 3\}$ $F = \{4, 5\}$

Coefficient matrix of original system of equations, and of associated homogenous system.

This matrix will be the subject of further analysis, rather than the systems of equations.

 $\begin{bmatrix} 2 & -3 & 1 & -6 \\ 4 & 1 & 2 & 9 \\ 3 & 1 & 1 & 8 \end{bmatrix}$

Matrix brought to reduced row-echelon form:

$\left[1 \right]$	0	0	2]
0	1	0	3
0	0	1	-1

Analysis of the row-reduced matrix (Notation RREFA [64]):

$$r = 3$$
 $D = \{1, 2, 3\}$ $F = \{4\}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the

coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

 $\left\langle \left\{ \begin{bmatrix} -2\\ -3\\ 1\\ 1 \end{bmatrix} \right\} \right\rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

 $\left\langle \left\{ \begin{bmatrix} 2\\4\\3 \end{bmatrix}, \begin{bmatrix} -3\\1\\1 \end{bmatrix}, \begin{bmatrix} 1\\2\\1 \end{bmatrix} \right\} \right\rangle$

The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$L = \begin{bmatrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

$\left\langle \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\} \right\rangle$

Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

$\left\langle \left\{ \begin{bmatrix} 1\\0\\0\\2 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\3 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\-1 \end{bmatrix} \right\} \right\rangle$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 4 Rank: 3 Nullity: 1

Archetype D

Summary System with three equations, four variables. Consistent. Null space of coefficient matrix has dimension 2. Coefficient matrix identical to that of Archetype E, vector of constants is different.

A system of linear equations (Definition SLE [18]):

$$2x_1 + x_2 + 7x_3 - 7x_4 = 8$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = -12
x₁ + x₂ + 4x₃ - 5x₄ = 4

Some solutions to the system of linear equations (not necessarily exhaustive):

 $x_1 = 0, \quad x_2 = 1, \quad x_3 = 2, \quad x_4 = 1$

 $x_1 = 4$, $x_2 = 0$, $x_3 = 0$, $x_4 = 0$

$$x_1 = 7, \quad x_2 = 8, \quad x_3 = 1, \quad x_4 = 3$$

Augmented matrix of the linear system of equations (Definition AM [57]):

 $\begin{bmatrix} 2 & 1 & 7 & -7 & 8 \\ -3 & 4 & -5 & -6 & -12 \\ 1 & 1 & 4 & -5 & 4 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$\left[1 \right]$	0	3	-2	4
0	1	1	-3	0
0	0	0	0	0

Analysis of the augmented matrix (Notation RREFA [64]):

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3, 4, 5\}$

U Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the

pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$2x_1 + x_2 + 7x_3 - 7x_4 = 0$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = 0
x₁ + x₂ + 4x₃ - 5x₄ = 0

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

$$x_1 = 0, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$$

Version 2.99

$$x_1 = -3, \quad x_2 = -1, \quad x_3 = 1, \quad x_4 = 0$$

$$x_1 = 2, \quad x_2 = 3, \quad x_3 = 0, \quad x_4 = 1$$

$$x_1 = -1, \quad x_2 = 2, \quad x_3 = 1, \quad x_4 = 1$$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

 $\begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3, 4, 5\}$

Coefficient matrix of original system of equations, and of associated homogenous system.

This matrix will be the subject of further analysis, rather than the systems of equations.

 $\begin{bmatrix} 2 & 1 & 7 & -7 \\ -3 & 4 & -5 & -6 \\ 1 & 1 & 4 & -5 \end{bmatrix}$

Matrix brought to reduced row-echelon form:

 $\begin{bmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 2 $D = \{1, 2\}$ $F = \{3, 4\}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the

coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

$$\left\langle \left\{ \begin{bmatrix} -3\\-1\\1\\0 \end{bmatrix}, \begin{bmatrix} 2\\3\\0\\1 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

 $\left\langle \left\{ \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 1\\4\\1 \end{bmatrix} \right\} \right\rangle$

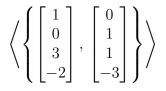
The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$L = \begin{bmatrix} 1 & \frac{1}{7} & -\frac{11}{7} \end{bmatrix}$ $\left\langle \left\{ \begin{bmatrix} \frac{11}{7} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{7} \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

$\left\langle \left\{ \begin{bmatrix} 1\\0\\\frac{7}{11} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{1}{11} \end{bmatrix} \right\} \right\rangle$

Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])



Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 4 Rank: 2 Nullity: 2

Archetype E

Summary System with three equations, four variables. Inconsistent. Null space of coefficient matrix has dimension 2. Coefficient matrix identical to that of Archetype D, constant vector is different.

A system of linear equations (Definition SLE [18]):

$$2x_1 + x_2 + 7x_3 - 7x_4 = 2$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = 3
x₁ + x₂ + 4x₃ - 5x₄ = 2

Some solutions to the system of linear equations (not necessarily exhaustive):

None. (Why?)

Augmented matrix of the linear system of equations (Definition AM [57]):

 $\begin{bmatrix} 2 & 1 & 7 & -7 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 1 & 1 & 4 & -5 & 2 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$\left[1 \right]$	0		-2	-
0	1	1	-3	0
0		0	0	1

Analysis of the augmented matrix (Notation RREFA [64]):

$$r = 3$$
 $D = \{1, 2, 5\}$ $F = \{3, 4\}$

U Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the

pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

Inconsistent system, no solutions exist.

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$2x_1 + x_2 + 7x_3 - 7x_4 = 0$$

-3x₁ + 4x₂ - 5x₃ - 6x₄ = 0
x₁ + x₂ + 4x₃ - 5x₄ = 0

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

 $x_1 = 0, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$

$$x_1 = 4, \quad x_2 = 13, \quad x_3 = 2, \quad x_4 = 5$$

Form the augmented matrix of the homogenous linear system, and use row operations

to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

$$\begin{bmatrix} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3, 4, 5\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.

$$\begin{bmatrix} 2 & 1 & 7 & -7 \\ -3 & 4 & -5 & -6 \\ 1 & 1 & 4 & -5 \end{bmatrix}$$

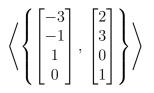
Matrix brought to reduced row-echelon form:

 $\begin{bmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Analysis of the row-reduced matrix (Notation RREFA [64]):

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3, 4\}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.



Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

 $\left\langle \left\{ \begin{bmatrix} 2\\-3\\1 \end{bmatrix}, \begin{bmatrix} 1\\4\\1 \end{bmatrix} \right\} \right\rangle$

The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

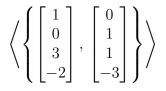
 $L = \begin{bmatrix} 1 & \frac{1}{7} & -\frac{11}{7} \end{bmatrix}$

$\left\langle \left\{ \begin{bmatrix} \frac{11}{7} \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -\frac{1}{7} \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

 $\left\langle \left\{ \begin{bmatrix} 1\\0\\\frac{7}{11} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{1}{11} \end{bmatrix} \right\} \right\rangle$

Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])



Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 4 Rank: 2 Nullity: 2

Archetype F

Summary System with four equations, four variables. Nonsingular coefficient matrix. Integer eigenvalues, one has "high" multiplicity.

A system of linear equations (Definition SLE [18]):

$$33x_1 - 16x_2 + 10x_3 - 2x_4 = -27$$

$$99x_1 - 47x_2 + 27x_3 - 7x_4 = -77$$

$$78x_1 - 36x_2 + 17x_3 - 6x_4 = -52$$

$$-9x_1 + 2x_2 + 3x_3 + 4x_4 = 5$$

Some solutions to the system of linear equations (not necessarily exhaustive):

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = -2, \quad x_4 = 4$$

Augmented matrix of the linear system of equations (Definition AM [57]):

 $\begin{bmatrix} 33 & -16 & 10 & -2 & -27 \\ 99 & -47 & 27 & -7 & -77 \\ 78 & -36 & 17 & -6 & -52 \\ -9 & 2 & 3 & 4 & 5 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

Analysis of the augmented matrix (Notation RREFA [64]):

$$r = 4$$
 $D = \{1, 2, 3, 4\}$ $F = \{5\}$

Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

$\begin{bmatrix} x_1 \end{bmatrix}$		[1]
x_2	_	2
x_3	_	-2
x_4		4

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$33x_1 - 16x_2 + 10x_3 - 2x_4 = 0$$

$$99x_1 - 47x_2 + 27x_3 - 7x_4 = 0$$

$$78x_1 - 36x_2 + 17x_3 - 6x_4 = 0$$

$$-9x_1 + 2x_2 + 3x_3 + 4x_4 = 0$$

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

 $x_1 = 0, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

$\left[1 \right]$	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 4$$
 $D = \{1, 2, 3, 4\}$ $F = \{5\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.

 $\begin{bmatrix} 33 & -16 & 10 & -2 \\ 99 & -47 & 27 & -7 \\ 78 & -36 & 17 & -6 \\ -9 & 2 & 3 & 4 \end{bmatrix}$

Matrix brought to reduced row-echelon form:

1	0	0	0]
0	1	0	0
0	0	1	0
0	0	0	1

Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 4 $D = \{1, 2, 3, 4\}$ $F = \{ \}$

 \Box Matrix (coefficient matrix) is nonsingular or singular? (Theorem NMRRI [176]) at the same time, examine the size of the set F above.Notice that this property does not apply to matrices that are not square.

Nonsingular.

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

$\langle \{ \} \rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

$$\left\langle \left\{ \begin{bmatrix} 33\\99\\78\\-9 \end{bmatrix}, \begin{bmatrix} -16\\-47\\-36\\2 \end{bmatrix}, \begin{bmatrix} 10\\27\\17\\3 \end{bmatrix}, \begin{bmatrix} -2\\-7\\-6\\4 \end{bmatrix} \right\} \right\rangle$$

The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

 $L = \begin{bmatrix} \\ \\ \\ \\ \\ \\ \\ \end{bmatrix} \begin{pmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \end{pmatrix} \right\rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example

CSROI [587], this yields a linearly independent set of vectors that span the column space.

$$\left\langle \left\{ \begin{bmatrix} 1\\0\\0\\0\end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\end{bmatrix} \right\} \right\rangle$$

Bow space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

$$\left\langle \left\{ \begin{bmatrix} 1\\0\\0\\0\end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\end{bmatrix} \right\} \right\rangle$$

Inverse matrix, if it exists. The inverse is not defined for matrices that are not square, and if the matrix is square, then the matrix must be nonsingular. (Definition MI [506], Theorem NI [542])

$$\begin{bmatrix} -\left(\frac{86}{3}\right) & \frac{38}{3} & -\left(\frac{11}{3}\right) & \frac{7}{3} \\ -\left(\frac{129}{2}\right) & \frac{86}{3} & -\left(\frac{17}{2}\right) & \frac{31}{6} \\ -13 & 6 & -2 & 1 \\ -\left(\frac{45}{2}\right) & \frac{29}{3} & -\left(\frac{5}{2}\right) & \frac{13}{6} \end{bmatrix}$$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 4 Rank: 4 Nullity: 0

Determinant of the matrix, which is only defined for square matrices. The matrix is nonsingular if and only if the determinant is nonzero (Theorem SMZD [935]). (Product of all eigenvalues?)

Determinant = -18

Eigenvalues, and bases for eigenspaces. (Definition EEM [945], Definition EM [963])

$$\lambda = -1 \qquad \qquad \mathcal{E}_{F}(-1) = \left\langle \left\{ \begin{bmatrix} 1\\2\\0\\1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 2 \qquad \qquad \mathcal{E}_{F}(2) = \left\langle \left\{ \begin{bmatrix} 2\\5\\2\\1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 3 \qquad \qquad \mathcal{E}_{F}(3) = \left\langle \left\{ \begin{bmatrix} 1\\1\\0\\7 \end{bmatrix}, \begin{bmatrix} 17\\45\\21\\0 \end{bmatrix} \right\} \right\rangle$$

Geometric and algebraic multiplicities. (Definition GME [969]Definition AME [968])

$$\gamma_F(-1) = 1 \qquad \qquad \alpha_F(-1) = 1$$

$$\gamma_F(2) = 1 \qquad \qquad \alpha_F(2) = 1$$

$$\gamma_F(3) = 2 \qquad \qquad \alpha_F(3) = 2$$

Diagonalizable? (Definition DZM [1037]) Yes, full eigenspaces, Theorem DMFE [1044].

The diagonalization. (Theorem DC [1039])

$$\begin{bmatrix} 12 & -5 & 1 & -1 \\ -39 & 18 & -7 & 3 \\ \frac{27}{7} & -\frac{13}{7} & \frac{6}{7} & -\frac{1}{7} \\ \frac{26}{7} & -\frac{12}{7} & \frac{5}{7} & -\frac{2}{7} \end{bmatrix} \begin{bmatrix} 33 & -16 & 10 & -2 \\ 99 & -47 & 27 & -7 \\ 78 & -36 & 17 & -6 \\ -9 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 17 \\ 2 & 5 & 1 & 45 \\ 0 & 2 & 0 & 21 \\ 1 & 1 & 7 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Archetype G

Summary System with five equations, two variables. Consistent. Null space of coefficient matrix has dimension 0. Coefficient matrix identical to that of Archetype H, constant vector is different.

A system of linear equations (Definition SLE [18]):

$$2x_1 + 3x_2 = 6$$

$$-x_1 + 4x_2 = -14$$

$$3x_1 + 10x_2 = -2$$

$$3x_1 - x_2 = 20$$

$$6x_1 + 9x_2 = 18$$

Some solutions to the system of linear equations (not necessarily exhaustive):

$$x_1 = 6, \quad x_2 = -2$$

Augmented matrix of the linear system of equations (Definition AM [57]):

 $\begin{bmatrix} 2 & 3 & 6 \\ -1 & 4 & -14 \\ 3 & 10 & -2 \\ 3 & -1 & 20 \\ 6 & 9 & 18 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

1	0	6]
0	1	-2
0	$\overline{0}$	0
0	0	0
0	0	0

Analysis of the augmented matrix (Notation RREFA [64]):

r = 2 $D = \{1, 2\}$ $F = \{3\}$

Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

 $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$2x_1 + 3x_2 = 0$$

$$-x_1 + 4x_2 = 0$$

$$3x_1 + 10x_2 = 0$$

$$3x_1 - x_2 = 0$$

$$6x_1 + 9x_2 = 0$$

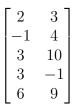
Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive): $x_1 = 0, \quad x_2 = 0$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

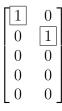
Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.



Matrix brought to reduced row-echelon form:



Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 2 $D = \{1, 2\}$ $F = \{ \}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

 $\langle \{ \} \rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

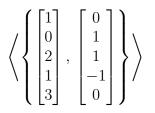
$$\left\langle \left\{ \begin{bmatrix} 2\\-1\\3\\3\\6 \end{bmatrix}, \begin{bmatrix} 3\\4\\10\\-1\\9 \end{bmatrix} \right\} \right\rangle$$

The column space of the matrix, as it arises from the extended echelon form of the

matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 1 - \frac{1}{3} & \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$
$$\left\langle \left\{ \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.



Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

 $\left\langle \left\{ \begin{bmatrix} 1\\ 0 \end{bmatrix}, \begin{bmatrix} 0\\ 1 \end{bmatrix} \right\} \right\rangle$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 2 Rank: 2 Nullity: 0

Archetype H

Summary System with five equations, two variables. Inconsistent, overdetermined. Null space of coefficient matrix has dimension 0. Coefficient matrix identical to that of Archetype G, constant vector is different.

A system of linear equations (Definition SLE [18]):

$$2x_{1} + 3x_{2} = 5$$

$$-x_{1} + 4x_{2} = 6$$

$$3x_{1} + 10x_{2} = 2$$

$$3x_{1} - x_{2} = -3$$

$$6x_{1} + 9x_{2} = 3$$

Some solutions to the system of linear equations (not necessarily exhaustive):

None. (Why?)

Augmented matrix of the linear system of equations (Definition AM [57]):

 $\begin{bmatrix} 2 & 3 & 5 \\ -1 & 4 & 6 \\ 3 & 10 & 2 \\ 3 & -1 & -1 \\ 6 & 9 & 3 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$\left[1 \right]$	0	0]
0	1	0
0	0	1
0	0	0
0	0	0

Analysis of the augmented matrix (Notation RREFA [64]):

r = 3 $D = \{1, 2, 3\}$ $F = \{\}$

Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

Inconsistent system, no solutions exist.

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$2x_1 + 3x_2 = 0$$

$$-x_1 + 4x_2 = 0$$

$$3x_1 + 10x_2 = 0$$

$$3x_1 - x_2 = 0$$

$$6x_1 + 9x_2 = 0$$

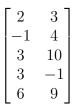
Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive): $x_1 = 0, \quad x_2 = 0$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

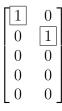
Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2$$
 $D = \{1, 2\}$ $F = \{3\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.



Matrix brought to reduced row-echelon form:



Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 2 $D = \{1, 2\}$ $F = \{ \}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

 $\langle \{ \} \rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

$$\left\langle \left\{ \begin{bmatrix} 2\\-1\\3\\3\\6 \end{bmatrix}, \begin{bmatrix} 3\\4\\10\\-1\\9 \end{bmatrix} \right\} \right\rangle$$

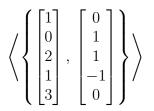
The column space of the matrix, as it arises from the extended echelon form of the

matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

L = []

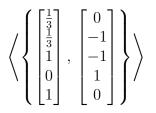
 $\left\langle \left\{ \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\} \right\rangle$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.



[] The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 1 - \frac{1}{3} \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$



Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

 $\left\langle \left\{ \begin{bmatrix} 1\\ 0 \end{bmatrix}, \begin{bmatrix} 0\\ 1 \end{bmatrix} \right\} \right\rangle$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 2 Rank: 2 Nullity: 0

Archetype I

Summary System with four equations, seven variables. Consistent. Null space of coefficient matrix has dimension 4.

A system of linear equations (Definition SLE [18]):

$$x_1 + 4x_2 - x_4 + 7x_6 - 9x_7 = 3$$

$$2x_1 + 8x_2 - x_3 + 3x_4 + 9x_5 - 13x_6 + 7x_7 = 9$$

$$2x_3 - 3x_4 - 4x_5 + 12x_6 - 8x_7 = 1$$

$$-x_1 - 4x_2 + 2x_3 + 4x_4 + 8x_5 - 31x_6 + 37x_7 = 4$$

Some solutions to the system of linear equations (not necessarily exhaustive):

 $x_1 = -25, x_2 = 4, x_3 = 22, x_4 = 29, x_5 = 1, x_6 = 2, x_7 = -3$ $x_1 = -7, x_2 = 5, x_3 = 7, x_4 = 15, x_5 = -4, x_6 = 2, x_7 = 1$

$$x_1 = 4, x_2 = 0, x_3 = 2, x_4 = 1, x_5 = 0, x_6 = 0, x_7 = 0$$

Augmented matrix of the linear system of equations (Definition AM [57]): $\begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 & 3 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 & 9 \end{bmatrix}$

 $\begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 & 3 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 & 9 \\ 0 & 0 & 2 & -3 & -4 & 12 & -8 & 1 \\ -1 & -4 & 2 & 4 & 8 & -31 & 37 & 4 \end{bmatrix}$

Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

1	4	0	0	2	1	-3	4
0	0	1	0	1	$ \begin{array}{r} 1 \\ -3 \\ -6 \\ 0 \end{array} $	5	2
0	0	0	1	2	-6	6	1
0	0	0	0	0	0	0	0

Analysis of the augmented matrix (Notation RREFA [64]):

r = 3 $D = \{1, 3, 4\}$ $F = \{2, 5, 6, 7, 8\}$

Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.

$\begin{bmatrix} x_1 \end{bmatrix}$]	$\left\lceil 4 \right\rceil$		$\begin{bmatrix} -4 \end{bmatrix}$		$\left[-2\right]$		[-1]		[3]
x_2		0		1		0		0		0
x_3		2		0		-1		3		-5
x_4	=	1	$+x_{2}$	0	$+x_{5}$	-2	$+x_{6}$	6	$+x_{7}$	-6
x_5		0		0		1		0		0
x_6		0		0		0		1		0
$\lfloor x_7 \rfloor$		0		0		0		0		1

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$x_1 + 4x_2 - x_4 + 7x_6 - 9x_7 = 0$$

$$2x_1 + 8x_2 - x_3 + 3x_4 + 9x_5 - 13x_6 + 7x_7 = 0$$

$$2x_3 - 3x_4 - 4x_5 + 12x_6 - 8x_7 = 0$$
$$-x_1 - 4x_2 + 2x_3 + 4x_4 + 8x_5 - 31x_6 + 37x_7 = 0$$

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

$$x_1 = 0, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 0, x_6 = 0, x_7 = 0$$

$$x_1 = 3, x_2 = 0, x_3 = -5, x_4 = -6, x_5 = 0, x_6 = 0, x_7 = 1$$

- $x_1 = -1, x_2 = 0, x_3 = 3, x_4 = 6, x_5 = 0, x_6 = 1, x_7 = 0$
- $x_1 = -2, x_2 = 0, x_3 = -1, x_4 = -2, x_5 = 1, x_6 = 0, x_7 = 0$
- $x_1 = -4, x_2 = 1, x_3 = 0, x_4 = 0, x_5 = 0, x_6 = 0, x_7 = 0$
- $x_1 = -4, x_2 = 1, x_3 = -3, x_4 = -2, x_5 = 1, x_6 = 1, x_7 = 1$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain

zeros:

$\left[1 \right]$	4	0	0	2	1	-3	0
0	0	1	0	1	-3	5	0
0	0	0	1	2	-6	6	0
0	0	0	0	0	$ \begin{array}{r} 1 \\ -3 \\ -6 \\ 0 \end{array} $	0	0

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 3$$
 $D = \{1, 3, 4\}$ $F = \{2, 5, 6, 7, 8\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.

[1]	4	0	-1	0	7	-97
2	8	-1	3	9	-13	7
0	0	2	-3	-4	12	-8
$\lfloor -1 \rfloor$	-4	2	4	8	$7 \\ -13 \\ 12 \\ -31$	37

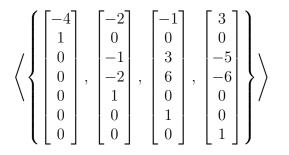
Matrix brought to reduced row-echelon form:

 $\begin{bmatrix} 1 & 4 & 0 & 0 & 2 & 1 & -3 \\ 0 & 0 & 1 & 0 & 1 & -3 & 5 \\ 0 & 0 & 0 & 1 & 2 & -6 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

Analysis of the row-reduced matrix (Notation RREFA [64]):

$$r = 3$$
 $D = \{1, 3, 4\}$ $F = \{2, 5, 6, 7\}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.



Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

$$\left\langle \left\{ \begin{bmatrix} 1\\2\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\-1\\2\\2 \end{bmatrix}, \begin{bmatrix} -1\\3\\-3\\4 \end{bmatrix} \right\} \right\rangle$$

The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m,

the matrix L has no rows and the column space is all of \mathbb{C}^m .

$$L = \begin{bmatrix} 1 & -\frac{12}{31} & -\frac{13}{31} & \frac{7}{31} \end{bmatrix}$$
$$\left\langle \left\{ \begin{bmatrix} -\frac{7}{31} \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} \frac{13}{31} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{12}{31} \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

$$\left\langle \left\{ \begin{bmatrix} 1\\0\\0\\-\frac{31}{7} \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\\frac{12}{7} \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\\frac{13}{7} \end{bmatrix} \right\} \right\rangle$$

Bow space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (The-

orem BRS [583])

 $\left\langle \left\{ \begin{array}{ccccccc} -4 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 2 & 1 & 2 & 1 \\ 1 & -3 & -6 & \end{array} \right\} \right\rangle$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 7 Rank: 3 Nullity: 4

Archetype J

Summary System with six equations, nine variables. Consistent. Null space of coefficient matrix has dimension 5.

A system of linear equations (Definition SLE [18]):

$$x_{1} + 2x_{2} - 2x_{3} + 9x_{4} + 3x_{5} - 5x_{6} - 2x_{7} + x_{8} + 27x_{9} = -5$$

$$2x_{1} + 4x_{2} + 3x_{3} + 4x_{4} - x_{5} + 4x_{6} + 10x_{7} + 2x_{8} - 23x_{9} = 18$$

$$x_{1} + 2x_{2} + x_{3} + 3x_{4} + x_{5} + x_{6} + 5x_{7} + 2x_{8} - 7x_{9} = 6$$

$$2x_{1} + 4x_{2} + 3x_{3} + 4x_{4} - 7x_{5} + 2x_{6} + 4x_{7} - 11x_{9} = 20$$

$$x_{1} + 2x_{2} + 5x_{4} + 2x_{5} - 4x_{6} + 3x_{7} + 8x_{8} + 13x_{9} = -4$$

$$-3x_{1} - 6x_{2} - x_{3} - 13x_{4} + 2x_{5} - 5x_{6} - 4x_{7} + 13x_{8} + 10x_{9} = -29$$

Some solutions to the system of linear equations (not necessarily exhaustive): $x_1 = 6, x_2 = 0, x_3 = -1, x_4 = 0, x_5 = -1, x_6 = 2, x_7 = 0, x_8 = 0, x_9 = 0$

$$x_1 = 4, x_2 = 1, x_3 = -1, x_4 = 0, x_5 = -1, x_6 = 2, x_7 = 0, x_8 = 0, x_9 = 0$$

 $x_1 = -17, x_2 = 7, x_3 = 3, x_4 = 2, x_5 = -1, x_6 = 14, x_7 = -1, x_8 = 3, x_9 = 2$
 $x_1 = -11, x_2 = -6, x_3 = 1, x_4 = 5, x_5 = -4, x_6 = 7, x_7 = 3, x_8 = 1, x_9 = 1$

Augmented matrix of the linear system of equations (Definition AM [57]):

$$\begin{bmatrix} 1 & 2 & -2 & 9 & 3 & -5 & -2 & 1 & 27 & -5 \\ 2 & 4 & 3 & 4 & -1 & 4 & 10 & 2 & -23 & 18 \\ 1 & 2 & 1 & 3 & 1 & 1 & 5 & 2 & -7 & 6 \\ 2 & 4 & 3 & 4 & -7 & 2 & 4 & 0 & -11 & 20 \\ 1 & 2 & 0 & 5 & 2 & -4 & 3 & 8 & 13 & -4 \\ -3 & -6 & -1 & -13 & 2 & -5 & -4 & 13 & 10 & -29 \end{bmatrix}$$

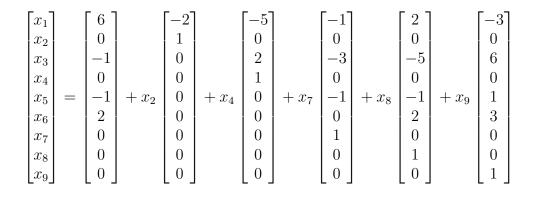
Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$\lceil 1 \rceil$	2	0	5	0	0	1	-2	3	6]
0	0	1	-2	0	0	3	5	-6	-1
0	0	0	0	1	0	1	1	-1	-1
0	0	0	0	0	1	0	-2	-3	2
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Analysis of the augmented matrix (Notation RREFA [64]):

r = 4 $D = \{1, 3, 5, 6\}$ $F = \{2, 4, 7, 8, 9, 10\}$

Vector form of the solution set to the system of equations (Theorem VFSLS [244]). Notice the relationship between the free variables and the set F above. Also, notice the pattern of 0's and 1's in the entries of the vectors corresponding to elements of the set F for the larger examples.



Given a system of equations we can always build a new, related, homogeneous system (Definition HS [146]) by converting the constant terms to zeros and retaining the coefficients of the variables. Properties of this new system will have precise relationships with various properties of the original system.

$$\begin{aligned} x_1 + 2x_2 - 2x_3 + 9x_4 + 3x_5 - 5x_6 - 2x_7 + x_8 + 27x_9 &= 0\\ 2x_1 + 4x_2 + 3x_3 + 4x_4 - x_5 + 4x_6 + 10x_7 + 2x_8 - 23x_9 &= 0\\ x_1 + 2x_2 + x_3 + 3x_4 + x_5 + x_6 + 5x_7 + 2x_8 - 7x_9 &= 0\\ 2x_1 + 4x_2 + 3x_3 + 4x_4 - 7x_5 + 2x_6 + 4x_7 - 11x_9 &= 0\\ x_1 + 2x_2 + 5x_4 + 2x_5 - 4x_6 + 3x_7 + 8x_8 + 13x_9 &= 0\\ -3x_1 - 6x_2 - x_3 - 13x_4 + 2x_5 - 5x_6 - 4x_7 + 13x_8 + 10x_9 &= 0 \end{aligned}$$

 $\begin{array}{|c|c|c|c|c|c|c|} \hline & \text{Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):} \\ & x_1 = 0, \, x_2 = 0, \, x_3 = 0, \, x_4 = 0, \, x_5 = 0, \, x_6 = 0, \, x_7 = 0, \, x_8 = 0, \, x_9 = 0 \\ \hline & x_1 = -2, \, x_2 = 1, \, x_3 = 0, \, x_4 = 0, \, x_5 = 0, \, x_6 = 0, \, x_7 = 0, \, x_8 = 0, \, x_9 = 0 \\ \hline & x_1 = -23, \, x_2 = 7, \, x_3 = 4, \, x_4 = 2, \, x_5 = 0, \, x_6 = 12, \, x_7 = -1, \, x_8 = 3, \, x_9 = 2 \\ \hline & x_1 = -17, \, x_2 = -6, \, x_3 = 2, \, x_4 = 5, \, x_5 = -3, \, x_6 = 5, \, x_7 = 3, \, x_8 = 1, \, x_9 = 1 \\ \end{array}$

Form the augmented matrix of the homogenous linear system, and use row operations to convert to reduced row-echelon form. Notice how the entries of the final column remain zeros:

Analysis of the augmented matrix for the homogenous system (Notation RREFA [64]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 4$$
 $D = \{1, 3, 5, 6\}$ $F = \{2, 4, 7, 8, 9, 10\}$

Coefficient matrix of original system of equations, and of associated homogenous system. This matrix will be the subject of further analysis, rather than the systems of equations.

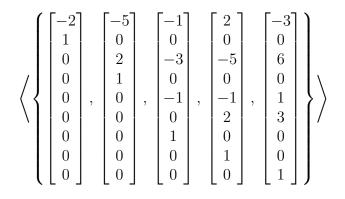
$$\begin{bmatrix} 1 & 2 & -2 & 9 & 3 & -5 & -2 & 1 & 27 \\ 2 & 4 & 3 & 4 & -1 & 4 & 10 & 2 & -23 \\ 1 & 2 & 1 & 3 & 1 & 1 & 5 & 2 & -7 \\ 2 & 4 & 3 & 4 & -7 & 2 & 4 & 0 & -11 \\ 1 & 2 & 0 & 5 & 2 & -4 & 3 & 8 & 13 \\ -3 & -6 & -1 & -13 & 2 & -5 & -4 & 13 & 10 \end{bmatrix}$$

Matrix brought to reduced row-echelon form:

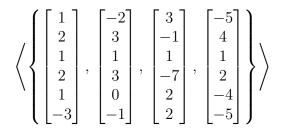
Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 4 $D = \{1, 3, 5, 6\}$ $F = \{2, 4, 7, 8, 9\}$

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

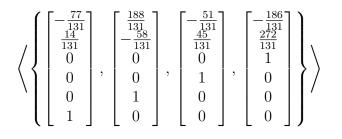


Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])



The column space of the matrix, as it arises from the extended echelon form of the matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$$L = \begin{bmatrix} 1 & 0 & \frac{186}{131} & \frac{51}{131} & -\frac{188}{131} & \frac{77}{131} \\ 0 & 1 & -\frac{272}{131} & -\frac{45}{131} & \frac{58}{131} & -\frac{14}{131} \end{bmatrix}$$

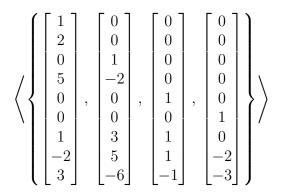


Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.

$$\left\langle \left\{ \begin{bmatrix} 1\\0\\0\\0\\-1\\-\frac{12}{7}\end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0\\-\frac{11}{2}\\-\frac{94}{7}\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\10\\22\end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\\\frac{3}{2}\\3\end{bmatrix} \right\} \right\rangle$$

Row space of the matrix, expressed as a span of a set of linearly independent vectors,

obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])



Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 9 Rank: 4 Nullity: 5

Archetype K

Summary Square matrix of size 5. Nonsingular. 3 distinct eigenvalues, 2 of multiplicity 2.

A matrix:

1 0	18	24	24	-12
12	-2	-6	0	-18
-30	-21	-23	-30	39
27	30	36	37	-30
18	24	30	30	-20

Matrix brought to reduced row-echelon form:

 $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 5 $D = \{1, 2, 3, 4, 5\}$ $F = \{\}$

 \square Matrix (coefficient matrix) is nonsingular or singular? (Theorem NMRRI [176]) at the same time, examine the size of the set F above.Notice that this property does not apply to matrices that are not square.

Nonsingular.

This is the null space of the matrix. The set of vectors used in the span construction is a linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

 $\langle \{ \} \rangle$

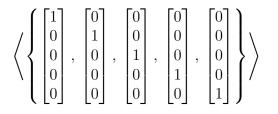
Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

$$\left\langle \left\{ \begin{bmatrix} 10\\12\\-30\\27\\18 \end{bmatrix}, \begin{bmatrix} 18\\-2\\-21\\30\\24 \end{bmatrix}, \begin{bmatrix} 24\\-6\\-23\\36\\30 \end{bmatrix}, \begin{bmatrix} 24\\0\\-30\\37\\30 \end{bmatrix}, \begin{bmatrix} -12\\-18\\39\\-30\\-20 \end{bmatrix} \right\} \right\rangle$$

The column space of the matrix, as it arises from the extended echelon form of the

matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.



Bow space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

Inverse matrix, if it exists. The inverse is not defined for matrices that are not square, and if the matrix is square, then the matrix must be nonsingular. (Definition MI [506], Theorem NI [542])

$$\begin{bmatrix} 1 & -\left(\frac{9}{4}\right) & -\left(\frac{3}{2}\right) & 3 & -6\\ \frac{21}{2} & \frac{43}{4} & \frac{21}{2} & 9 & -9\\ -15 & -\left(\frac{21}{2}\right) & -11 & -15 & \frac{39}{2}\\ 9 & \frac{15}{4} & \frac{9}{2} & 10 & -15\\ \frac{9}{2} & \frac{3}{4} & \frac{3}{2} & 6 & -\left(\frac{19}{2}\right) \end{bmatrix}$$

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 5 Rank: 5 Nullity: 0

Determinant of the matrix, which is only defined for square matrices. The matrix is nonsingular if and only if the determinant is nonzero (Theorem SMZD [935]). (Product of all eigenvalues?)

Determinant = 16

Eigenvalues, and bases for eigenspaces. (Definition EEM [945], Definition EM [963])

$$\lambda = -2 \qquad \qquad \mathcal{E}_{K}(-2) = \left\langle \left\{ \begin{bmatrix} 2\\-2\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} -1\\2\\-2\\1\\0 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 1 \qquad \qquad \mathcal{E}_{K}(1) = \left\langle \left\{ \begin{bmatrix} 4\\-10\\7\\0\\2 \end{bmatrix}, \begin{bmatrix} -4\\18\\-17\\5\\0 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 4 \qquad \qquad \mathcal{E}_{K}(4) = \left\langle \left\{ \begin{bmatrix} 1\\-1\\0\\1\\1 \end{bmatrix} \right\} \right\rangle$$

Geometric and algebraic multiplicities. (Definition GME [969]Definition AME [968])

$$\gamma_K(-2) = 2 \qquad \qquad \alpha_K(-2) = 2$$

$$\gamma_K (1) = 2 \qquad \qquad \alpha_K (1) = 2$$

$$\gamma_K (4) = 1 \qquad \qquad \alpha_K (4) = 1$$

Diagonalizable? (Definition DZM [1037])

Yes, full eigenspaces, Theorem DMFE [1044].

The diagonalization. (Theorem DC [1039])

$$\begin{bmatrix} -4 & -3 & -4 & -6 & 7 \\ -7 & -5 & -6 & -8 & 10 \\ 1 & -1 & -1 & 1 & -3 \\ 1 & 0 & 0 & 1 & -2 \\ 2 & 5 & 6 & 4 & 0 \end{bmatrix} \begin{bmatrix} 10 & 18 & 24 & 24 & -12 \\ 12 & -2 & -6 & 0 & -18 \\ -30 & -21 & -23 & -30 & 39 \\ 27 & 30 & 36 & 37 & -30 \\ 18 & 24 & 30 & 30 & -20 \end{bmatrix} \begin{bmatrix} 2 & -1 & 4 & -4 & 1 \\ -2 & 2 & -10 & 18 & -1 \\ 1 & -2 & 7 & -17 & 0 \\ 0 & 1 & 0 & 5 & 1 \\ 1 & 0 & 2 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} -2 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

Archetype L

Summary Square matrix of size 5. Singular, nullity 2. 2 distinct eigenvalues, each of "high" multiplicity.

A matrix:

$$\begin{bmatrix} -2 & -1 & -2 & -4 & 4 \\ -6 & -5 & -4 & -4 & 6 \\ 10 & 7 & 7 & 10 & -13 \\ -7 & -5 & -6 & -9 & 10 \\ -4 & -3 & -4 & -6 & 6 \end{bmatrix}$$

Matrix brought to reduced row-echelon form:

Analysis of the row-reduced matrix (Notation RREFA [64]):

r = 5 $D = \{1, 2, 3\}$ $F = \{4, 5\}$

 \square Matrix (coefficient matrix) is nonsingular or singular? (Theorem NMRRI [176]) at the same time, examine the size of the set F above.Notice that this property does not apply to matrices that are not square.

Singular.

This is the null space of the matrix. The set of vectors used in the span construction is a

linearly independent set of column vectors that spans the null space of the matrix (Theorem SSNS [283], Theorem BNS [334]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [244]) to see these vectors arise.

$$\left\langle \left\{ \begin{bmatrix} -1\\2\\-2\\1\\0 \end{bmatrix}, \begin{bmatrix} 2\\-2\\1\\0\\1 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors that are also columns of the matrix. These columns have indices that form the set D above. (Theorem BCS [569])

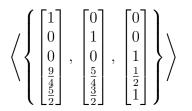
$$\left\langle \left\{ \begin{bmatrix} -2\\ -6\\ 10\\ -7\\ -4 \end{bmatrix}, \begin{bmatrix} -1\\ -5\\ 7\\ -5\\ -3 \end{bmatrix}, \begin{bmatrix} -2\\ -4\\ 7\\ -6\\ -4 \end{bmatrix} \right\} \right\rangle$$

The column space of the matrix, as it arises from the extended echelon form of the

matrix. The matrix L is computed as described in Definition EEF [619]. This is followed by the column space described by a set of linearly independent vectors that span the null space of L, computed as according to Theorem FS [625] and Theorem BNS [334]. When r = m, the matrix L has no rows and the column space is all of \mathbb{C}^m .

$$L = \begin{bmatrix} 1 & 0 & -2 & -6 & 5 \\ 0 & 1 & 4 & 10 & -9 \end{bmatrix}$$
$$\left\langle \left\{ \begin{bmatrix} -5 \\ 9 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 6 \\ -10 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\} \right\rangle$$

Column space of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose of the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining nonzero rows as column vectors. By Theorem CSRST [586] and Theorem BRS [583], and in the style of Example CSROI [587], this yields a linearly independent set of vectors that span the column space.



Row space of the matrix, expressed as a span of a set of linearly independent vectors, obtained from the nonzero rows of the equivalent matrix in reduced row-echelon form. (Theorem BRS [583])

$$\left\langle \left\{ \begin{bmatrix} 1\\0\\0\\1\\-2\end{bmatrix}, \begin{bmatrix} 0\\1\\0\\-2\\2\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\2\\-1\end{bmatrix} \right\} \right\rangle$$

Inverse matrix, if it exists. The inverse is not defined for matrices that are not square, and if the matrix is square, then the matrix must be nonsingular. (Definition MI [506], Theorem NI [542])

Subspace dimensions associated with the matrix. (Definition NOM [829], Definition ROM [829]) Verify Theorem RPNC [831]

Matrix columns: 5 Rank: 3 Nullity: 2

Determinant of the matrix, which is only defined for square matrices. The matrix is nonsingular if and only if the determinant is nonzero (Theorem SMZD [935]). (Product of all eigenvalues?)

Determinant = 0

Eigenvalues, and bases for eigenspaces. (Definition EEM [945], Definition EM [963])

$$\lambda = -1 \qquad \qquad \mathcal{E}_L \left(-1\right) = \left\langle \left\{ \begin{bmatrix} -5\\9\\0\\0\\1 \end{bmatrix}, \begin{bmatrix} 6\\-10\\0\\1\\0 \end{bmatrix}, \begin{bmatrix} 2\\-4\\1\\0\\0 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 0 \qquad \qquad \mathcal{E}_L(0) = \left\langle \left\{ \begin{bmatrix} 2\\-2\\1\\0\\1 \end{bmatrix}, \begin{bmatrix} -1\\2\\-2\\1\\0 \end{bmatrix} \right\} \right\rangle$$

Geometric and algebraic multiplicities. (Definition GME [969]Definition AME [968])

$$\gamma_L (-1) = 3 \qquad \qquad \alpha_L (-1) = 3 \gamma_L (0) = 2 \qquad \qquad \alpha_L (0) = 2$$



Yes, full eigenspaces, Theorem DMFE [1044].

The diagonalization. (Theorem DC [1039])

Archetype M

Summary Linear transformation with bigger domain than codomain, so it is guaranteed to not be injective. Happens to not be surjective.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^5 \to \mathbb{C}^3, \quad T\left(\begin{bmatrix} x_1\\x_2\\x_3\\x_4\\x_5 \end{bmatrix} \right) = \begin{bmatrix} x_1 + 2x_2 + 3x_3 + 4x_4 + 4x_5\\3x_1 + x_2 + 4x_3 - 3x_4 + 7x_5\\x_1 - x_2 - 5x_4 + x_5 \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

$$\left\{ \begin{bmatrix} -2\\ -1\\ 0\\ 0\\ 1 \end{bmatrix}, \begin{bmatrix} 2\\ -3\\ 0\\ 1\\ 0 \end{bmatrix}, \begin{bmatrix} -1\\ -1\\ 1\\ 0\\ 0 \end{bmatrix} \right\}$$

Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. Also, since the rank can not exceed 3, we are guaranteed to have a nullity of at least 2, just from checking dimensions of the domain and the codomain. In particular, verify that

$$T\left(\begin{bmatrix}1\\2\\-1\\4\\5\end{bmatrix}\right) = \begin{bmatrix}38\\24\\-16\end{bmatrix} \qquad T\left(\begin{bmatrix}0\\-3\\0\\5\\6\end{bmatrix}\right) = \begin{bmatrix}38\\24\\-16\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} 0\\ -3\\ 0\\ 5\\ 6 \end{bmatrix} = \begin{bmatrix} 1\\ 2\\ -1\\ 4\\ 5 \end{bmatrix} + \begin{bmatrix} -1\\ -5\\ 1\\ 1\\ 1\\ 1 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} -1\\ -5\\ 1\\ 1\\ 1 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} 1\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\1\\-1 \end{bmatrix}, \begin{bmatrix} 3\\4\\0 \end{bmatrix}, \begin{bmatrix} 4\\-3\\-5 \end{bmatrix}, \begin{bmatrix} 4\\7\\1 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1\\0\\-\frac{4}{5} \end{bmatrix}, \begin{bmatrix} 0\\1\\\frac{3}{5} \end{bmatrix} \right\}$$

Surjective: No. (Definition SLT [1169])

Notice that the range is not all of \mathbb{C}^3 since it has dimension 2, not 3. In particular, verify that $\begin{bmatrix} 3\\4\\5 \end{bmatrix} \notin \mathcal{R}(T)$, by setting the output equal to this vector and seeing that the resulting

system of linear equations has no solution, i.e. is inconsistent. So the preimage, $T^{-1} \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}$,

is empty. This alone is sufficient to see that the linear transformation is not onto.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

	Domain dimension:	5	Rank: 2	Nullity: 3
--	-------------------	---	---------	------------

Invertible: No.

Not injective or surjective.

Matrix representation (Theorem MLTCV [1089]):

$$T: \mathbb{C}^5 \to \mathbb{C}^3, \quad T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} 1 & 2 & 3 & 4 & 4 \\ 3 & 1 & 4 & -3 & 7 \\ 1 & -1 & 0 & -5 & 1 \end{bmatrix}$$

Archetype N

Summary Linear transformation with domain larger than its codomain, so it is guaranteed to not be injective. Happens to be onto.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^5 \to \mathbb{C}^3, \quad T\left(\begin{bmatrix} x_1\\x_2\\x_3\\x_4\\x_5 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 + x_2 + 3x_3 - 4x_4 + 5x_5\\x_1 - 2x_2 + 3x_3 - 9x_4 + 3x_5\\3x_1 + 4x_3 - 6x_4 + 5x_5 \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

$$\left\{ \begin{bmatrix} 1\\-1\\-2\\0\\1 \end{bmatrix}, \begin{bmatrix} -2\\-1\\3\\1\\0 \end{bmatrix} \right\}$$

Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. Also, since the rank can not exceed 3, we are guaranteed to have a nullity of at least 2, just from checking dimensions of the domain and the codomain. In particular, verify that

$$T\left(\begin{bmatrix}-3\\1\\-2\\-3\\1\end{bmatrix}\right) = \begin{bmatrix}6\\19\\6\end{bmatrix} \qquad T\left(\begin{bmatrix}-4\\-4\\-2\\-1\\4\end{bmatrix}\right) = \begin{bmatrix}6\\19\\6\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} -4\\ -4\\ -2\\ -2\\ -1\\ 4 \end{bmatrix} = \begin{bmatrix} -3\\ 1\\ -2\\ -3\\ 1 \end{bmatrix} + \begin{bmatrix} -1\\ -5\\ 0\\ 2\\ 3 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} -1\\ -5\\ 0\\ 2\\ 3 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} 2\\1\\3 \end{bmatrix}, \begin{bmatrix} 1\\-2\\0 \end{bmatrix}, \begin{bmatrix} 3\\3\\4 \end{bmatrix}, \begin{bmatrix} -4\\-9\\-6 \end{bmatrix}, \begin{bmatrix} 5\\3\\5 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\}$$

Surjective: Yes. (Definition SLT [1169])

Notice that the basis for the range above is the standard basis for \mathbb{C}^3 . So the range is all of \mathbb{C}^3 and thus the linear transformation is surjective.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

```
Domain dimension: 5 Rank: 3 Nullity: 2
```

Invertible: No.

Not surjective, and the relative sizes of the domain and codomain mean the linear transformation cannot be injective. (Theorem ILTIS [1221])

Matrix representation (Theorem MLTCV [1089]):

$$T: \mathbb{C}^5 \to \mathbb{C}^3, \quad T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} 2 & 1 & 3 & -4 & 5 \\ 1 & -2 & 3 & -9 & 3 \\ 3 & 0 & 4 & -6 & 5 \end{bmatrix}$$

Archetype O

Summary Linear transformation with a domain smaller than the codomain, so it is guaranteed to not be onto. Happens to not be one-to-one.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^3 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} -x_1 + x_2 - 3x_3 \\ -x_1 + 2x_2 - 4x_3 \\ x_1 + x_2 + x_3 \\ 2x_1 + 3x_2 + x_3 \\ x_1 + 2x_3 \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

 $\left\{ \begin{bmatrix} -2\\1\\1 \end{bmatrix} \right\}$

Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. Also, since the rank can not exceed 3, we are guaranteed to have a nullity of at least 2, just from checking dimensions of the domain and the codomain. In particular, verify that

$$T\left(\begin{bmatrix}5\\-1\\3\end{bmatrix}\right) = \begin{bmatrix}-15\\-19\\7\\10\\11\end{bmatrix} \qquad T\left(\begin{bmatrix}1\\1\\5\end{bmatrix}\right) = \begin{bmatrix}-15\\-19\\7\\10\\11\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} 1\\1\\5 \end{bmatrix} = \begin{bmatrix} 5\\-1\\3 \end{bmatrix} + \begin{bmatrix} -4\\2\\2 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} -4\\2\\2 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} -1\\ -1\\ 1\\ 2\\ 1 \end{bmatrix}, \begin{bmatrix} 1\\ 2\\ 1\\ 3\\ 0 \end{bmatrix}, \begin{bmatrix} -3\\ -4\\ 1\\ 1\\ 1\\ 2 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps

un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1\\0\\-3\\-7\\-2 \end{bmatrix}, \begin{bmatrix} 0\\1\\2\\5\\1 \end{bmatrix} \right\}$$

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension: 3 Rank: 2 Nullity: 1



The dimension of the range is 2, and the codomain (\mathbb{C}^5) has dimension 5. So the transformation is not onto. Notice too that since the domain \mathbb{C}^3 has dimension 3, it is impossible for the range to have a dimension greater than 3, and no matter what the actual definition of the function, it cannot possibly be onto. To be more precise, verify that $\begin{bmatrix} 2\\3\\1\\1\\1 \end{bmatrix} \notin \mathcal{R}(T)$, by setting the output equal to this vector and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent. So the preimage, $T^{-1}\begin{pmatrix} 2\\3\\1\\1\\1 \end{pmatrix}$, is empty. This alone is sufficient to see that the linear transformation is not onto.

transformation is not onto

Invertible: No.

Not injective, and the relative dimensions of the domain and codomain prohibit any possibility of being surjective.

Matrix representation (Theorem MLTCV [1089]):

$$T: \mathbb{C}^3 \to \mathbb{C}^5, \quad T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} -1 & 1 & -3\\ -1 & 2 & -4\\ 1 & 1 & 1\\ 2 & 3 & 1\\ 1 & 0 & 2 \end{bmatrix}$$

Archetype P

Summary Linear transformation with a domain smaller that its codomain, so it is guaranteed to not be surjective. Happens to be injective.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^{3} \to \mathbb{C}^{5}, \quad T\left(\begin{bmatrix}x_{1}\\x_{2}\\x_{3}\end{bmatrix}\right) = \begin{bmatrix}-x_{1} + x_{2} + x_{3}\\-x_{1} + 2x_{2} + 2x_{3}\\x_{1} + x_{2} + 3x_{3}\\2x_{1} + 3x_{2} + x_{3}\\-2x_{1} + x_{2} + 3x_{3}\end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

Injective: Yes. (Definition ILT [1128])

Since $\mathcal{K}(T) = \{\mathbf{0}\}$, Theorem KILT [1144] tells us that T is injective.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} -1\\ -1\\ 1\\ 1\\ 2\\ -2 \end{bmatrix}, \begin{bmatrix} 1\\ 2\\ 1\\ 3\\ 1\\ 1 \end{bmatrix}, \begin{bmatrix} 1\\ 2\\ 3\\ 1\\ 3\\ 1\\ 3 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps

un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1\\0\\0\\-10\\6 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\7\\-3 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\-1\\1 \end{bmatrix} \right\}$$

Surjective: No. (Definition SLT [1169])

The dimension of the range is 3, and the codomain (\mathbb{C}^5) has dimension 5. So the transformation is not surjective. Notice too that since the domain \mathbb{C}^3 has dimension 3, it is impossible for the range to have a dimension greater than 3, and no matter what the actual definition of the function, it cannot possibly be surjective in this situation.

To be more precise, verify that $\begin{bmatrix} 2\\1\\-3\\2\\6 \end{bmatrix} \notin \mathcal{R}(T)$, by setting the output equal to this vector and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent.

So the preimage, $T^{-1}\begin{pmatrix} 2\\1\\-3\\2\\6 \end{bmatrix}$, is empty. This alone is sufficient to see that the linear transformation is not onto.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

```
Domain dimension: 3 Rank: 3 Nullity: 0
```

```
Invertible: No.
```

The relative dimensions of the domain and codomain prohibit any possibility of being surjective, so apply Theorem ILTIS [1221].

Matrix representation (Theorem MLTCV [1089]):

$$T: \mathbb{C}^3 \to \mathbb{C}^5, \quad T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} -1 & 1 & 1 \\ -1 & 2 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 1 \\ -2 & 1 & 3 \end{bmatrix}$$

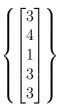
Archetype Q

Summary Linear transformation with equal-sized domain and codomain, so it has the potential to be invertible, but in this case is not. Neither injective nor surjective. Diagonalizable, though.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1\\x_2\\x_3\\x_4\\x_5 \end{bmatrix} \right) = \begin{bmatrix} -2x_1 + 3x_2 + 3x_3 - 6x_4 + 3x_5\\-16x_1 + 9x_2 + 12x_3 - 28x_4 + 28x_5\\-19x_1 + 7x_2 + 14x_3 - 32x_4 + 37x_5\\-21x_1 + 9x_2 + 15x_3 - 35x_4 + 39x_5\\-9x_1 + 5x_2 + 7x_3 - 16x_4 + 16x_5 \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])



Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. Also, since the rank can not exceed 3, we are guaranteed to have a nullity of at least 2, just from checking dimensions of the domain and the codomain. In particular, verify that

$$T\left(\begin{bmatrix}1\\3\\-1\\2\\4\end{bmatrix}\right) = \begin{bmatrix}4\\55\\72\\77\\31\end{bmatrix} \qquad T\left(\begin{bmatrix}4\\7\\0\\5\\7\end{bmatrix}\right) = \begin{bmatrix}4\\55\\72\\77\\31\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} 4\\7\\0\\5\\7\end{bmatrix} = \begin{bmatrix} 1\\3\\-1\\+\\1\\3\\3\end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} 3\\4\\1\\3\\3 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} -2\\ -16\\ -19\\ -21\\ -9 \end{bmatrix}, \begin{bmatrix} 3\\ 9\\ 7\\ 9\\ 5 \end{bmatrix}, \begin{bmatrix} 3\\ 12\\ 14\\ 15\\ 7 \end{bmatrix}, \begin{bmatrix} -6\\ -28\\ -32\\ -35\\ -16 \end{bmatrix}, \begin{bmatrix} 3\\ 28\\ 37\\ 39\\ 16 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1\\0\\0\\0\\0\\1 \end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\\0\\-1 \end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\\2 \end{bmatrix} \right\}$$

Surjective: No. (Definition SLT [1169])

The dimension of the range is 4, and the codomain (\mathbb{C}^5) has dimension 5. So $\mathcal{R}(T) \neq \mathbb{C}^5$ and by Theorem RSLT [1185] the transformation is not surjective. To be more precise, verify that $\begin{bmatrix} -1\\ 2\\ 3\\ -1\\ 4 \end{bmatrix} \notin \mathcal{R}(T)$, by setting the output equal to this vector and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent. So the preimage, $T^{-1}\begin{pmatrix} \begin{bmatrix} -1\\ 2\\ 3\\ -1\\ 4 \end{bmatrix} \end{pmatrix}$, is empty. This alone is sufficient to see that the linear transformation is not onto.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Invertible: No.

Neither injective nor surjective. Notice that since the domain and codomain have the same

dimension, either the transformation is both onto and one-to-one (making it invertible) or else it is both not onto and not one-to-one (as in this case) by Theorem RPNDD [1235].

Matrix representation (Theorem MLTCV [1089]):

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} -2 & 3 & 3 & -6 & 3\\ -16 & 9 & 12 & -28 & 28\\ -19 & 7 & 14 & -32 & 37\\ -21 & 9 & 15 & -35 & 39\\ -9 & 5 & 7 & -16 & 16 \end{bmatrix}$$

Eigenvalues and eigenvectors (Definition EELT [1356], Theorem EER [1384]):

$$\lambda = -1 \qquad \qquad \mathcal{E}_T \left(-1 \right) = \left\langle \left\{ \begin{bmatrix} 0\\2\\3\\1\\1 \end{bmatrix} \right\} \right\rangle$$

$$\lambda = 0 \qquad \qquad \mathcal{E}_{T}(0) = \left\langle \left\{ \begin{bmatrix} 3\\4\\1\\3\\3 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 1 \qquad \qquad \mathcal{E}_{T}(1) = \left\langle \left\{ \begin{bmatrix} 5\\3\\0\\0\\2 \end{bmatrix}, \begin{bmatrix} -3\\1\\0\\2\\0 \end{bmatrix}, \begin{bmatrix} 1\\-1\\2\\0\\0 \end{bmatrix} \right\} \right\rangle$$

Evaluate the linear transformation with each of these eigenvectors as an interesting check.

A diagonal matrix representation relative to a basis of eigenvectors, B.

$$B = \left\{ \begin{bmatrix} 0\\2\\3\\3\\1 \end{bmatrix}, \begin{bmatrix} 3\\4\\1\\3\\3\\3 \end{bmatrix}, \begin{bmatrix} 5\\3\\0\\0\\2 \end{bmatrix}, \begin{bmatrix} -3\\1\\0\\2\\0 \end{bmatrix}, \begin{bmatrix} 1\\-1\\2\\0\\0\\2 \end{bmatrix} \right\}$$

Archetype R

Summary Linear transformation with equal-sized domain and codomain. Injective, surjective, invertible, diagonalizable, the works.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} -65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5 \\ 36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5 \\ -44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5 \\ 34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5 \\ 12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5 \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

Injective: Yes. (Definition ILT [1128])

Since the kernel is trivial Theorem KILT [1144] tells us that the linear transformation is injective.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} -65\\36\\-44\\34\\12 \end{bmatrix}, \begin{bmatrix} 128\\-73\\88\\-68\\-24 \end{bmatrix}, \begin{bmatrix} 10\\-1\\5\\-3\\-1 \end{bmatrix}, \begin{bmatrix} -262\\151\\-180\\140\\49 \end{bmatrix}, \begin{bmatrix} 40\\-16\\24\\-18\\-5 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps

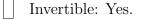
un-coordinatizing. A basis for the range is:

Surjective: Yes. (Definition SLT [1169])

A basis for the range is the standard basis of \mathbb{C}^5 , so $\mathcal{R}(T) = \mathbb{C}^5$ and Theorem RSLT [1185] tells us T is surjective. Or, the dimension of the range is 5, and the codomain (\mathbb{C}^5) has dimension 5. So the transformation is surjective.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension: 5 Rank: 5 Nullity: 0



Both injective and surjective (Theorem ILTIS [1221]). Notice that since the domain and codomain have the same dimension, either the transformation is both injective and surjective (making it invertible, as in this case) or else it is both not injective and not surjective.

Matrix representation (Theorem MLTCV [1089]):

$$T: \mathbb{C}^5 \to \mathbb{C}^5, \quad T(\mathbf{x}) = A\mathbf{x}, \quad A = \begin{bmatrix} -65 & 128 & 10 & -262 & 40 \\ 36 & -73 & -1 & 151 & -16 \\ -44 & 88 & 5 & -180 & 24 \\ 34 & -68 & -3 & 140 & -18 \\ 12 & -24 & -1 & 49 & -5 \end{bmatrix}$$

The inverse linear transformation (Definition IVLT [1214]):

$$T^{-1}: \mathbb{C}^5 \to \mathbb{C}^5, \quad T^{-1} \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \right) = \begin{bmatrix} -47x_1 + 92x_2 + x_3 - 181x_4 - 14x_5 \\ 27x_1 - 55x_2 + \frac{7}{2}x_3 + \frac{221}{2}x_4 + 11x_5 \\ -32x_1 + 64x_2 - x_3 - 126x_4 - 12x_5 \\ 25x_1 - 50x_2 + \frac{3}{2}x_3 + \frac{199}{2}x_4 + 9x_5 \\ 9x_1 - 18x_2 + \frac{1}{2}x_3 + \frac{71}{2}x_4 + 4x_5 \end{bmatrix}$$

Verify that $T^{-1}(T(\mathbf{x})) = \mathbf{x}$ and $T(T^{-1}(\mathbf{x})) = \mathbf{x}$, and notice that the representations of the transformation and its inverse are matrix inverses (Theorem IMR [1322], Definition MI

[506]).

Eigenvalues and eigenvectors (Definition EELT [1356], Theorem EER [1384]):

$$\lambda = -1 \qquad \qquad \mathcal{E}_{T}(-1) = \left\langle \left\{ \begin{bmatrix} -57\\0\\-18\\14\\5 \end{bmatrix}, \begin{bmatrix} 2\\1\\0\\0\\0\\0 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 1 \qquad \qquad \mathcal{E}_{T}(1) = \left\langle \left\{ \begin{bmatrix} -10\\-5\\-6\\0\\1 \end{bmatrix}, \begin{bmatrix} 2\\3\\1\\1\\0 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 2 \qquad \qquad \mathcal{E}_{T}(2) = \left\langle \left\{ \begin{bmatrix} -6\\3\\-4\\3\\1 \end{bmatrix} \right\} \right\rangle$$

. -

- -

Evaluate the linear transformation with each of these eigenvectors as an interesting check.

A diagonal matrix representation relative to a basis of eigenvectors, *B*.

$$B = \left\{ \begin{bmatrix} -57\\0\\-18\\14\\5 \end{bmatrix}, \begin{bmatrix} 2\\1\\0\\0\\0\\0 \end{bmatrix}, \begin{bmatrix} -10\\-5\\-6\\0\\1 \end{bmatrix}, \begin{bmatrix} 2\\3\\1\\1\\0\\0 \end{bmatrix}, \begin{bmatrix} -6\\3\\-4\\3\\1\\1 \end{bmatrix} \right\}$$
$$M_{B,B}^{T} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0\\0 & -1 & 0 & 0 & 0\\0 & 0 & 1 & 0 & 0\\0 & 0 & 0 & 1 & 0\\0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

Archetype S

Summary Domain is column vectors, codomain is matrices. Domain is dimension 3 and codomain is dimension 4. Not injective, not surjective.

A linear transformation: (Definition LT [1072])

$$T: \mathbb{C}^3 \to M_{22}, \quad T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix} \right) = \begin{bmatrix} a-b & 2a+2b+c \\ 3a+b+c & -2a-6b-2c \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

$$\left\{ \begin{bmatrix} -1\\ -1\\ 4 \end{bmatrix} \right\}$$

Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. Also, since the rank can not exceed 3, we are guaranteed to have a nullity of at least 1, just from checking dimensions of the domain and the codomain. In particular, verify that

$$T\left(\begin{bmatrix}2\\1\\3\end{bmatrix}\right) = \begin{bmatrix}1 & 9\\10 & -16\end{bmatrix} \qquad T\left(\begin{bmatrix}0\\-1\\11\end{bmatrix}\right) = \begin{bmatrix}1 & 9\\10 & -16\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} 0\\-1\\11 \end{bmatrix} = \begin{bmatrix} 2\\1\\3 \end{bmatrix} + \begin{bmatrix} -2\\-2\\8 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} -2\\ -2\\ 8 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} 1 & 2 \\ 3 & -2 \end{bmatrix}, \begin{bmatrix} -1 & 2 \\ 1 & -6 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & -2 \end{bmatrix} \right\}$$

Surjective: No. (Definition SLT [1169])

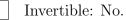
The dimension of the range is 2, and the codomain (M_{22}) has dimension 4. So the trans-

formation is not surjective. Notice too that since the domain \mathbb{C}^3 has dimension 3, it is impossible for the range to have a dimension greater than 3, and no matter what the actual definition of the function, it cannot possibly be surjective in this situation.

To be more precise, verify that $\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \notin \mathcal{R}(T)$, by setting the output of T equal to this matrix and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent. So the preimage, $T^{-1}\left(\begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \right)$, is empty. This alone is sufficient to see that the linear transformation is not onto.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

```
Domain dimension: 3 Rank: 2 Nullity: 1
```



Not injective (Theorem ILTIS [1221]), and the relative dimensions of the domain and codomain prohibit any possibility of being surjective.

Matrix representation (Definition MR [1286]):

$$B = \left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 0\\1\\0 \end{bmatrix}, \begin{bmatrix} 0\\0\\1 \end{bmatrix} \right\}$$
$$C = \left\{ \begin{bmatrix} 1&0\\0&0 \end{bmatrix}, \begin{bmatrix} 0&1\\0&0 \end{bmatrix}, \begin{bmatrix} 0&0\\1&0 \end{bmatrix}, \begin{bmatrix} 0&0\\0&1 \end{bmatrix} \right\}$$
$$M_{B,C}^{T} = \begin{bmatrix} 1&-1&0\\2&2&1\\3&1&1\\-2&-6&-2 \end{bmatrix}$$

Archetype T

Summary Domain and codomain are polynomials. Domain has dimension 5, while codomain has dimension 6. Is injective, can't be surjective.

A linear transformation: (Definition LT [1072])

 $T: P_4 \rightarrow P_5, \quad T(p(x)) = (x-2)p(x)$

A basis for the null space of the linear transformation: (Definition KLT [1138])

{ }

Injective: Yes. (Definition ILT [1128])

Since the kernel is trivial Theorem KILT [1144] tells us that the linear transformation is

injective.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\{x-2, x^2-2x, x^3-2x^2, x^4-2x^3, x^5-2x^4, x^6-2x^5\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{-\frac{1}{32}x^5+1, \ -\frac{1}{16}x^5+x, \ -\frac{1}{8}x^5+x^2, \ -\frac{1}{4}x^5+x^3, \ -\frac{1}{2}x^5+x^4\right\}$$

Surjective: No. (Definition SLT [1169])

The dimension of the range is 5, and the codomain (P_5) has dimension 6. So the transformation is not surjective. Notice too that since the domain P_4 has dimension 5, it is impossible for the range to have a dimension greater than 5, and no matter what the actual definition of the function, it cannot possibly be surjective in this situation.

To be more precise, verify that $1 + x + x^2 + x^3 + x^4 \notin \mathcal{R}(T)$, by setting the output equal to this vector and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent. So the preimage, $T^{-1}(1 + x + x^2 + x^3 + x^4)$, is nonempty. This alone is sufficient to see that the linear transformation is not onto.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension: 5 Rank: 5 Nullity: 0

Invertible: No.

The relative dimensions of the domain and codomain prohibit any possibility of being surjective, so apply Theorem ILTIS [1221].



Matrix representation (Definition MR [1286]):

$$B = \left\{1, \, x, \, x^2, \, x^3, \, x^4\right\}$$

$$C = \left\{ 1, x, x^2, x^3, x^4, x^5 \right\}$$
$$M_{B,C}^T = \begin{bmatrix} -2 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Archetype U

Summary Domain is matrices, codomain is column vectors. Domain has dimension 6, while codomain has dimension 4. Can't be injective, is surjective.

A linear transformation: (Definition LT [1072])

$$T: M_{23} \to \mathbb{C}^4, \quad T\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a+2b+12c-3d+e+6f \\ 2a-b-c+d-11f \\ a+b+7c+2d+e-3f \\ a+2b+12c+5e-5f \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

$$\left\{ \begin{bmatrix} 3 & -4 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} -2 & -5 & 1 \\ 0 & 0 & 0 \end{bmatrix} \right\}$$

Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. Also, since the rank can not exceed 4, we are guaranteed to have a nullity of at least 2, just from checking dimensions of the domain and the codomain. In particular, verify that

$$T\left(\begin{bmatrix}1 & 10 & -2\\3 & -1 & 1\end{bmatrix}\right) = \begin{bmatrix}-7\\-14\\-1\\-13\end{bmatrix} \qquad T\left(\begin{bmatrix}5 & -3 & -1\\5 & 3 & 3\end{bmatrix}\right) = \begin{bmatrix}-7\\-14\\-1\\-13\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} 5 & -3 & -1 \\ 5 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 10 & -2 \\ 3 & -1 & 1 \end{bmatrix} + \begin{bmatrix} 4 & -13 & 1 \\ 2 & 4 & 2 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} 4 & -13 & 1 \\ 2 & 4 & 2 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\begin{bmatrix} 1\\2\\1\\1\\1 \end{bmatrix}, \begin{bmatrix} 2\\-1\\1\\2\\2\\0 \end{bmatrix}, \begin{bmatrix} 12\\-1\\-1\\7\\12\\2\\0 \end{bmatrix}, \begin{bmatrix} -3\\1\\2\\0\\-1\\1\\5 \end{bmatrix}, \begin{bmatrix} 1\\0\\1\\5\\-1\\1\\-3\\-5 \end{bmatrix}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

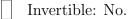
$$\left\{ \begin{bmatrix} 1\\0\\0\\0\end{bmatrix}, \begin{bmatrix} 0\\1\\0\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\1\\0\end{bmatrix}, \begin{bmatrix} 0\\0\\0\\1\end{bmatrix} \right\}$$

Surjective: Yes. (Definition SLT [1169])

A basis for the range is the standard basis of \mathbb{C}^4 , so $\mathcal{R}(T) = \mathbb{C}^4$ and Theorem RSLT [1185] tells us T is surjective. Or, the dimension of the range is 4, and the codomain (\mathbb{C}^4) has dimension 4. So the transformation is surjective.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension: 6 Rank: 4 Nullity: 2



The relative sizes of the domain and codomain mean the linear transformation cannot be injective. (Theorem ILTIS [1221])

Matrix representation (Definition MR [1286]):

$$B = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$
$$C = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$
$$M_{B,C}^{T} = \begin{bmatrix} 1 & 2 & 12 & -3 & 1 & 6 \\ 2 & -1 & -1 & 1 & 0 & -11 \\ 1 & 1 & 7 & 2 & 1 & -3 \\ 1 & 2 & 12 & 0 & 5 & -5 \end{bmatrix}$$

Archetype V

Summary Domain is polynomials, codomain is matrices. Domain and codomain both have dimension 4. Injective, surjective, invertible. Square matrix representation, but domain and codomain are unequal, so no eigenvalue information.

A linear transformation: (Definition LT [1072])

$$T: P_3 \to M_{22}, \quad T\left(a+bx+cx^2+dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

{ }

Injective: Yes. (Definition ILT [1128])

Since the kernel is trivial Theorem KILT [1144] tells us that the linear transformation is injective.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Surjective: Yes. (Definition SLT [1169])

A basis for the range is the standard basis of M_{22} , so $\mathcal{R}(T) = M_{22}$ and Theorem RSLT [1185] tells us T is surjective. Or, the dimension of the range is 4, and the codomain (M_{22}) has dimension 4. So the transformation is surjective.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension: 4 Rank: 4 Nullity: 0

Invertible: Yes.

Both injective and surjective (Theorem ILTIS [1221]). Notice that since the domain and codomain have the same dimension, either the transformation is both injective and surjective (making it invertible, as in this case) or else it is both not injective and not surjective.

Matrix representation (Definition MR [1286]):

$$B = \{1, x, x^{2}, x^{3}\}$$

$$C = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

$$M_{B,C}^{T} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

Since invertible, the inverse linear transformation. (Definition IVLT [1214])

$$T^{-1}: M_{22} \to P_3, \quad T^{-1}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a-c-d) + (c+d)x + \frac{1}{2}(a-b-c-d)x^2 + cx^3$$

Archetype W

Summary Domain is polynomials, codomain is polynomials. Domain and codomain both have dimension 3. Injective, surjective, invertible, 3 distinct eigenvalues, diagonalizable.

A linear transformation: (Definition LT [1072])

 $T: P_2 \to P_2, \quad T(a+bx+cx^2) = (19a+6b-4c) + (-24a-7b+4c) + (36a+12b-9c)$

A basis for the null space of the linear transformation: (Definition KLT [1138])

{ }

Injective: Yes. (Definition ILT [1128])

Since the kernel is trivial Theorem KILT [1144] tells us that the linear transformation is

injective.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\{19 - 24x + 36x^2, 6 - 7x + 12x^2, -4 + 4x - 9x^2\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\{1, x, x^2\}$$

Surjective: Yes. (Definition SLT [1169])

A basis for the range is the standard basis of \mathbb{C}^5 , so $\mathcal{R}(T) = \mathbb{C}^5$ and Theorem RSLT [1185] tells us T is surjective. Or, the dimension of the range is 5, and the codomain (\mathbb{C}^5) has dimension 5. So the transformation is surjective.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension:	3	Rank: 3	Nullity: 0
-------------------	---	---------	------------

Invertible: Yes.

Both injective and surjective (Theorem ILTIS [1221]). Notice that since the domain and codomain have the same dimension, either the transformation is both injective and surjective (making it invertible, as in this case) or else it is both not injective and not surjective.

Matrix representation (Definition MR [1286]):

$$B = \{1, x, x^2\}$$

$$C = \{1, x, x^2\}$$

$$M_{B,C}^T = \begin{bmatrix} 19 & 6 & -4\\ -24 & -7 & 4\\ 36 & 12 & -9 \end{bmatrix}$$

$$T^{-1}: P_2 \to P_2, \quad T^{-1}\left(a + bx + cx^2\right) = \left(-5a - 2b + \frac{4}{3}c\right) + \left(24a + 9b - \frac{20}{3}c\right)x + \left(12a + 4b - \frac{11}{3}c\right)x^2$$

Eigenvalues and eigenvectors (Definition EELT [1356], Theorem EER [1384]):

$$\lambda = -1 \qquad \qquad \mathcal{E}_T (-1) = \left\langle \left\{ 2x + 3x^2 \right\} \right\rangle \\ \lambda = 1 \qquad \qquad \mathcal{E}_T (1) = \left\langle \left\{ -1 + 3x \right\} \right\rangle \\ \lambda = 3 \qquad \qquad \mathcal{E}_T (3) = \left\langle \left\{ 1 - 2x + x^2 \right\} \right\rangle$$

Evaluate the linear transformation with each of these eigenvectors as an interesting check.

A diagonal matrix representation relative to a basis of eigenvectors, B.

$$B = \left\{ 2x + 3x^2, -1 + 3x, 1 - 2x + x^2 \right\}$$
$$M_{B,B}^T = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Version 2.99

Archetype X

Summary Domain and codomain are square matrices. Domain and codomain both have dimension 4. Not injective, not surjective, not invertible, 3 distinct eigenvalues, diagonalizable.

A linear transformation: (Definition LT [1072])

$$T: M_{22} \to M_{22}, \quad T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} -2a + 15b + 3c + 27d & 10b + 6c + 18d \\ a - 5b - 9d & -a - 4b - 5c - 8d \end{bmatrix}$$

A basis for the null space of the linear transformation: (Definition KLT [1138])

$$\left\{ \begin{bmatrix} -6 & -3 \\ 2 & 1 \end{bmatrix} \right\}$$

Injective: No. (Definition ILT [1128])

Since the kernel is nontrivial Theorem KILT [1144] tells us that the linear transformation is not injective. In particular, verify that

$$T\left(\begin{bmatrix}-2 & 0\\1 & -4\end{bmatrix}\right) = \begin{bmatrix}115 & 78\\-38 & -35\end{bmatrix} \qquad T\left(\begin{bmatrix}4 & 3\\-1 & 3\end{bmatrix}\right) = \begin{bmatrix}115 & 78\\-38 & -35\end{bmatrix}$$

This demonstration that T is not injective is constructed with the observation that

$$\begin{bmatrix} 4 & 3 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 1 & -4 \end{bmatrix} + \begin{bmatrix} 6 & 3 \\ -2 & -1 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} 6 & 3\\ -2 & -1 \end{bmatrix} \in \mathcal{K}(T)$$

so the vector \mathbf{z} effectively "does nothing" in the evaluation of T.

A basis for the range of the linear transformation: (Definition RLT [1178])

Evaluate the linear transformation on a standard basis to get a spanning set for the range (Theorem SSRLT [1188]):

$$\left\{ \begin{bmatrix} -2 & 0 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 15 & 10 \\ -5 & -4 \end{bmatrix}, \begin{bmatrix} 3 & 6 \\ 0 & -5 \end{bmatrix}, \begin{bmatrix} 27 & 18 \\ -9 & -8 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (Theorem ILTLI [1148]). This spanning set may be converted to a "nice" basis, by making the vectors the rows of a matrix (perhaps after using a vector reperesentation), row-reducing, and retaining the nonzero rows (Theorem BRS [583]), and perhaps un-coordinatizing. A basis for the range is:

$$\left\{ \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ \frac{1}{4} & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

Surjective: No. (Definition SLT [1169])

The dimension of the range is 3, and the codomain (M_{22}) has dimension 5. So $\mathcal{R}(T) \neq M_{22}$ and by Theorem RSLT [1185] the transformation is not surjective.

To be more precise, verify that $\begin{bmatrix} 2 & 4 \\ 3 & 1 \end{bmatrix} \notin \mathcal{R}(T)$, by setting the output of T equal to this matrix and seeing that the resulting system of linear equations has no solution, i.e. is

inconsistent. So the preimage, $T^{-1}\left(\begin{bmatrix}2 & 4\\ 3 & 1\end{bmatrix}\right)$, is empty. This alone is sufficient to see that the linear transformation is not onto.

Subspace dimensions associated with the linear transformation. Examine parallels with earlier results for matrices. Verify Theorem RPNDD [1235].

Domain dimension: 4 Rank: 3 Nullity: 1

Invertible: No.

Neither injective nor surjective (Theorem ILTIS [1221]). Notice that since the domain and codomain have the same dimension, either the transformation is both injective and surjective or else it is both not injective and not surjective (making it not invertible, as in this case).

Matrix representation (Definition MR [1286]):

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$
$$C = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

$$M_{B,C}^{T} = \begin{bmatrix} -2 & 15 & 3 & 27\\ 0 & 10 & 6 & 18\\ 1 & -5 & 0 & -9\\ -1 & -4 & -5 & -8 \end{bmatrix}$$

Eigenvalues and eigenvectors (Definition EELT [1356], Theorem EER [1384]):

$$\lambda = 0 \qquad \qquad \mathcal{E}_{T}(0) = \left\langle \left\{ \begin{bmatrix} -6 & -3\\ 2 & 1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 1 \qquad \qquad \mathcal{E}_{T}(1) = \left\langle \left\{ \begin{bmatrix} -7 & -2\\ 3 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -2\\ 0 & 1 \end{bmatrix} \right\} \right\rangle$$
$$\lambda = 3 \qquad \qquad \mathcal{E}_{T}(3) = \left\langle \left\{ \begin{bmatrix} -3 & -2\\ 1 & 1 \end{bmatrix} \right\} \right\rangle$$

Evaluate the linear transformation with each of these eigenvectors as an interesting check.

A diagonal matrix representation relative to a basis of eigenvectors, B.

$$B = \left\{ \begin{bmatrix} -6 & -3\\ 2 & 1 \end{bmatrix}, \begin{bmatrix} -7 & -2\\ 3 & 0 \end{bmatrix}, \begin{bmatrix} -1 & -2\\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -3 & -2\\ 1 & 1 \end{bmatrix} \right\}$$

$$M_{B,B}^{T} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Appendix GFDL GNU Free Documentation License

Version 1.2, November 2002 Copyright ©2000,2001,2002 Free Software Foundation, Inc.

59 Temple Place, Suite 330, Boston, MA 02111-1307 USA

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy

and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "Modified Version" of the Document means any work containing the Document or

a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition. The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions

of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.
- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties–for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided for under this License. Any other attempt to copy, modify, sublicense or distribute the Document is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

Copyright ©YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with...Texts." line with this:

with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software. Part T Topics

Section F Fields

DRAFT: THIS SECTION COMPLETE, BUT SUBJECT TO CHANGE

We have chosen to present introductory linear algebra in the Core (Part C [2]) using scalars from the set of complex numbers, \mathbb{C} . We could have instead chosen to use scalars from the set of real numbers, \mathbb{R} . This would have presented certain difficulties when we encountered characteristic polynomials with complex roots (Definition CP [960]) or when we needed to be sure every matrix had at least one eigenvalue (Theorem EMHE [953]). However, much of the basics would be unchanged. The definition of a vector space would not change, nor would the ideas of linear independence, spanning, or bases. Linear transformations would still behave the same and we would still obtain matrix representations, though our ideas about canonical forms would have to be adjusted slightly.

The real numbers and the complex numbers are both examples of what are called fields, and we can "do" linear algebra in just a bit more generality by letting our scalars take values from some unspecified field. So in this section we will describe exactly what constitutes a field, give some finite examples, and discuss another connection between fields and vector spaces. Vector spaces over finite fields are very important in certain applications, so this is partially background for other topics. As such, we will not prove every claim we make.

Subsection F Fields

Like a vector space, a field is a set along with two binary operations. The distinction is that both operations accept two elements of the set, and then produce a new element of the set. In a vector space we have two sets — the vectors and the scalars, and scalar multiplication mixes one of each to produce a vector. Here is the careful definition of a field.

Definition F Field

Suppose that F is a set upon which we have defined two operations: (1) **addition**, which combines two elements of F and is denoted by "+", and (2) **multiplication**, which combines two elements of F and is denoted by juxtaposition. Then F, along with the two operations, is a **field** if the following properties hold.

- ACF Additive Closure, Field If $\alpha, \beta \in F$, then $\alpha + \beta \in F$.
- MCF Multiplicative Closure, Field If $\alpha, \beta \in F$, then $\alpha\beta \in F$.
- CAF Commutativity of Addition, Field If $\alpha, \beta \in F$, then $\alpha + \beta = \beta + \alpha$.

- CMF Commutativity of Multiplication, Field If $\alpha, \beta \in F$, then $\alpha\beta = \beta\alpha$.
- AAF Additive Associativity, Field If $\alpha, \beta, \gamma \in F$, then $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$.
- MAF Multiplicative Associativity, Field If $\alpha, \beta, \gamma \in F$, then $\alpha(\beta\gamma) = (\alpha\beta)\gamma$.
- **DF Distributivity, Field** If $\alpha, \beta, \gamma \in F$, then $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.
- **ZF** Zero, Field There is an element, $0 \in F$, called **zero**, such that $\alpha + 0 = \alpha$ for all $\alpha \in F$.
- OF One, Field There is an element, $1 \in F$, called **one**, such that $\alpha(1) = \alpha$ for all $\alpha \in F$.
- AIF Additive Inverse, Field If $\alpha \in F$, then there exists $-\alpha \in F$ so that $\alpha + (-\alpha) = 0$.
- MIF Multiplicative Inverse, Field If $\alpha \in F$, $\alpha \neq 0$, then there exists $\frac{1}{\alpha} \in F$ so that $\alpha(\frac{1}{\alpha}) = 1$.

 \triangle

Mostly this definition says that all the good things you might expect, really do happen in a field. The one technicality is that the special element, 0, the additive identity element, does not have a multiplicative inverse. In other words, no dividing by zero.

This definition should remind you of Theorem PCNA [1596], and indeed, Theorem PCNA [1596] provides the justification for the statement that the complex numbers form a field. Another example of field is the set of rational numbers

$$\mathbb{Q} = \left\{ \left. \frac{p}{q} \right| p, \ q \text{ are integers}, \ q \neq 0 \right\}$$

Of course, the real numbers, \mathbb{R} , also form a field. It is this field that you probably studied for many years. You began studying the integers ("counting"), then the rationals ("fractions"), then the reals ("algebra"), along with some excursions in the complex numbers ("imaginary numbers"). So you should have seen three fields already in your previous studies.

Our first observation about fields is that we can go back to our definition of a vector space (Definition VS [659]) and replace every occurrence of \mathbb{C} by some general, unspecified field, F, and all our subsequent definitions and theorems are still true, so long as we avoid roots of polynomials (or equivalently, factoring polynomials). So if you consult more advanced texts on linear algebra, you will see this sort of approach. You might study some of the first theorems we proved about vector spaces in Subsection VS.VSP [674] and work through their proofs in the more general setting of an arbitrary field. This exercise should convince you that very little changes when we move from \mathbb{C} to an arbitrary field F. (See Exercise F.T10 [1843].)

Subsection FF Finite Fields

It may sound odd at first, but there exist finite fields, and even finite vector spaces. We will find certain of these important in subsequent applications, so we collect some ideas and properties here.

Definition IMP Integers Modulo a Prime

Suppose that p is a prime number. Let $\mathbb{Z}_p = \{0, 1, 2, \dots, p-1\}$. Add and multiply elements of \mathbb{Z}_p as integers, but whenever a result lies outside of the set \mathbb{Z}_p , find its remainder after division by p and replace the result by this remainder. \triangle

We have defined a set, and two binary operations. The result is a field.

Theorem FIMP Field of Integers Modulo a Prime

The set of integers modulo a prime p, \mathbb{Z}_p , is a field.

Example IM11

Integers mod 11

 \mathbb{Z}_{11} is a field by Theorem FIMP [1833]. Here we provide some sample calculations.

8+5=2 -8=3 5-9=7

$$5(7) = 2$$
 $\frac{1}{7} = 8$
 $\frac{6}{5} = 10$
 $2^5 = 10$
 $-1 = 10$
 $\frac{1}{0} = ?$

 \boxtimes

We can now "do" linear algebra using scalars from a finite field.

Example VSIM5

Vector space over integers mod 5

Let $(\mathbb{Z}_5)^3$ be the set of all column vectors of length 3 with entries from \mathbb{Z}_5 . Use \mathbb{Z}_5 as the set of scalars. Define addition and multiplication the usual way. We exhibit a few sample calculations.

$$\begin{bmatrix} 2\\3\\4 \end{bmatrix} + \begin{bmatrix} 4\\1\\3 \end{bmatrix} = \begin{bmatrix} 1\\4\\2 \end{bmatrix} \qquad \qquad 3\begin{bmatrix} 2\\0\\4 \end{bmatrix} = \begin{bmatrix} 1\\0\\2 \end{bmatrix}$$

We can, of course, build linear combinations, such as

$$2\begin{bmatrix}1\\3\\0\end{bmatrix} - 4\begin{bmatrix}2\\1\\1\end{bmatrix} + \begin{bmatrix}1\\2\\4\end{bmatrix} = \begin{bmatrix}0\\4\\0\end{bmatrix}$$

which almost looks like a relation of linear dependence. The set

$$\left\{ \begin{bmatrix} 1\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\2\\0 \end{bmatrix} \right\}$$

is linearly independent, while the set

$$\left\{ \begin{bmatrix} 1\\3\\1 \end{bmatrix}, \begin{bmatrix} 2\\2\\0 \end{bmatrix}, \begin{bmatrix} 4\\3\\2 \end{bmatrix} \right\}$$

is linearly dependent, as can be seen from the relation of linear dependence formed by the scalars $a_1 = 2$, $a_2 = 1$ and $a_3 = 4$. To find these scalars, one would take the same approach as Example LDS [317], but in performing row operations to solve a homogeneous system, you would need to take care that all scalar (field) operations are performed over \mathbb{Z}_5 , especially when multiplying a row by a scalar to make a leading entry equal to 1. One more observation about this example — the set

$$\left\{ \begin{bmatrix} 1\\0\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\0 \end{bmatrix}, \begin{bmatrix} 1\\1\\1 \end{bmatrix} \right\}$$

is a basis for $(\mathbb{Z}_5)^3$, since it is both linearly independent and spans $(\mathbb{Z}_5)^3$.

In applications to computer science or electrical engineering, \mathbb{Z}_2 is the most important field, since it can be used to describe the binary nature of logic, circuitry, communications

 \boxtimes

and their intertwined relationships. The vector space of column vectors with entries from \mathbb{Z}_2 , $(\mathbb{Z}_2)^n$, with scalars taken from \mathbb{Z}_2 is the natural extension of this idea. Notice that \mathbb{Z}_2 has the minimum number of elements to be a field, since any field must contain a zero and a one (Property ZF [1831], Property OF [1831]).

Example SM2Z7

Symmetric matrices of size 2 over \mathbb{Z}_7

We can employ the field of integers modulo a prime to build other examples of vector spaces with novel fields of scalars. Define

$$S_{22}\left(\mathbb{Z}_{7}\right) = \left\{ \begin{bmatrix} a & b \\ b & c \end{bmatrix} \middle| a, b, c \in \mathbb{Z}_{7} \right\}$$

which is the set of all 2×2 symmetric matrices with entries from \mathbb{Z}_7 . Use the field \mathbb{Z}_7 as the set of scalars, and define vector addition and scalar multiplication in the natural way. The result will be a vector space.

Notice that the field of scalars is finite, as is the vector space, since there are $7^3 = 343$ matrices in $S_{22}(\mathbb{Z}_7)$. The set

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is a basis, so dim $(S_{22}(\mathbb{Z}_7)) = 3.$

In a more advanced algebra course it is possible to prove that the number of elements in a finite field must be of the form p^n , where p is a prime. We can't go so far afield as to prove this here, but we can demonstrate an example.

 \boxtimes

Example FF8 Finite field of size 8

Define the set F as $F = \{a + bt + ct^2 | a, b, c \in \mathbb{Z}_2\}$. Add and multiply these quantities as polynomials in the variable t, but replace any occurrence of t^3 by t + 1.

This defines a set, and the two operations on elements of that set. Do not be concerned with what t "is," because it isn't. t is just a handy device that makes the example a field. We'll say a bit more about t when we finish. But first, some examples. Remember that 1 + 1 = 0 in \mathbb{Z}_2 . Addition is quite simple, for example,

$$(1+t+t^2) + (1+t^2) = (1+1) + (1+0)t + (1+1)t^2 = t$$

Multiplication gets more involved, for example,

$$(1+t+t^{2})(1+t^{2}) = 1+t^{2}+t+t^{3}+t^{2}+t^{4}$$
$$= 1+t+(1+1)t^{2}+t^{3}(1+t)$$
$$= 1+t+(1+t)(1+t)$$
$$= 1+t+1+t+t+t^{2}$$
$$= (1+1)+(1+1+1)t+t^{2}$$
$$= t+t^{2}$$

Every element has a multiplicative inverse (Property MIF [1831]). What is the inverse of

 $t + t^2$? Check that

$$(t + t^{2}) (1 + t) = t + t^{2} + t^{2} + t^{3}$$
$$= t + (1 + 1)t^{2} + (1 + t)$$
$$= t + 1 + t$$
$$= 1 + (1 + 1)t$$
$$= 1$$

So we can write $\frac{1}{t+t^2} = 1 + t$. So that you may experiment, we give you the complete addition and multiplication tables for this field. Addition is simple, while multiplication is more interesting, so verify a few entries of each table. Because of the commutativity of addition and multiplication (Property CAF [1830], Property CMF [1831]), we have just listed half of each table.

+	0	1	t	t^2	t+1	$t^{2} + t$	$t^2 + t + 1$	$t^2 + 1$
0	0	1	t	t^2	t+1	$t^2 + t$	$t^2 + t + 1$	$t^2 + 1$
1		0	t+1	$t^2 + 1$	t	$t^2 + t + 1$	$t^2 + t$	t^2
t			0	$t^{2} + t$	1	t^2	$t^2 + 1$	$t^2 + t + 1$
t^2				0	$t^2 + t + 1$	t	t+1	1
t+1					0	$t^2 + 1$	t^2	$t^{2} + t$
$t^2 + t$						0	1	t+1
$t^2 + t + 1$							0	t
$t^2 + 1$								0

•	0	1	t	t^2	t+1	$t^{2} + t$	$t^2 + t + 1$	$t^2 + 1$
0	0	0	0	0	0	0	0	0
1		1	t	t^2	t+1	$t^2 + t$	$t^2 + t + 1$	$t^2 + 1$
t			t^2	t+1	$t^{2} + t$	$t^2 + t + 1$	$t^2 + 1$	1
t^2				$t^2 + t$	$t^2 + t + 1$	$t^2 + 1$	1	t
t+1					$t^2 + 1$	1	t	t^2
$t^{2} + t$						t	t^2	t+1
$t^2 + t + 1$							1+t	$t^2 + t$
$t^2 + 1$								$t^2 + t + 1$

Note that every element of F is a linear combination (with scalars from \mathbb{Z}_2) of the polynomials 1, t, t^2 . So $B = \{1, t, t^2\}$ is a spanning set for F. Further, B is linearly independent since there is no nontrivial relation of linear dependence, and B is a basis. So dim (F) = 3. Of course, this paragraph presumes that F is also a vector space over \mathbb{Z}_2 (which it is).

The defining relation for t ($t^3 = t + 1$) in Example FF8 [1837] arises from the polynomial $t^3 + t + 1$, which has no factorization with coefficients from \mathbb{Z}_2 . This is an example of an **irreducible polynomial**, which involves considerable theory to fully understand. In the exercises, we provide you with a few more irreducible polynomials to experiment with. See the suggested readings if you would like to learn more.

Trivially, every field (finite or otherwise) is a vector space. Suppose we begin with a field F. From this we know F has two binary operations defined on it. We need to somehow create a vector space from F, in a general way. First we need a set of vectors. That'll be F. We also need a set of scalars. That'll be F as well. How do we define the addition of two

vectors? By the same rule that we use to add them when they are in the field. How do we define scalar multiplication? Since a scalar is an element of F, and a vector is an element of F, we can define scalar multiplication to be the same rule that we use to multiply the two elements as members of the field. With these definitions, F will be a vector space (Exercise F.T20 [1843]). This is something of a trivial situation, since the set of vectors and the set of scalars are identical. In particular, do not confuse this with Example FF8 [1837] where the set of vectors has eight elements, and the set of scalars has just two elements.

Further Reading

Robert J. McEliece, Finite Fields for Scientists and Engineers. Kluwer Academic Publishers, 1987.

Rudolpf Lidl, Harald Niederreiter, Introduction to Finite Fields and Their Applications, Revised Edition. Cambridge University Press, 1994.

Subsection EXC Exercises

C60 Consider the vector space $(\mathbb{Z}_5)^4$ composed of column vectors of size 4 with entries from \mathbb{Z}_5 . The matrix A is a square matrix composed of four such column vectors.

$$A = \begin{bmatrix} 3 & 3 & 0 & 3 \\ 1 & 2 & 3 & 0 \\ 1 & 1 & 0 & 2 \\ 4 & 2 & 2 & 1 \end{bmatrix}$$

Find the inverse of A. Use this to find a solution to $\mathcal{LS}(A, \mathbf{b})$ when

$$\mathbf{b} = \begin{bmatrix} 3\\ 3\\ 2\\ 0 \end{bmatrix}$$

Contributed by Robert Beezer Solution [1844]

M10 Suppose we relax the restriction in Definition IMP [1833] to allow p to not be a prime. Will the construction given still be a field? Is \mathbb{Z}_6 a field? Can you generalize? Contributed by Robert Beezer M40 Construct a finite field with 9 elements using the set

$$F = \{a + bt \mid a, b \in \mathbb{Z}_3\}$$

where t^2 is consistently replaced by 2t + 1 in any intermediate results obtained with polynomial multiplication. Compute the first nine powers of t (t^0 through t^8). Use this information to aid you in the construction of the multiplication table for this field. What is the multiplicative inverse of 2t?

Contributed by Robert Beezer

M45 Construct a finite field with 25 elements using the set

$$F = \{a + bt \mid a, b \in \mathbb{Z}_5\}$$

where t^2 is consistently replaced by t+3 in any intermediate results obtained with polynomial multiplication. Compute the first 25 powers of t (t^0 through t^{24}). Use this information to aid you in computing in this field. What is the multiplicative inverse of 2t? What is the multiplicative inverse of 4? What is the multiplicative inverse of 1 + 4t?

Find a basis for F as a vector space with \mathbb{Z}_5 used as the set of scalars. Contributed by Robert Beezer

M50 Construct a finite field with 16 elements using the set

$$F = \left\{ a + bt + ct^{2} + dt^{3} \middle| a, b, c, d \in \mathbb{Z}_{2} \right\}$$

where t^4 is consistently replaced by t+1 in any intermediate results obtained with polynomial multiplication. Compute the first 16 powers of t (t^0 through t^{15}). Consider the set $G = \{0, 1, t^5, t^{10}\}$. Then G will also be a finite field, a subfield of F. Construct the addition and multiplication tables for G. Notice that since both G and F are vector spaces over \mathbb{Z}_2 , and $G \subseteq F$, by Definition S [694], G is a subspace of F. Contributed by Robert Beezer

T10 Give a new proof of Theorem ZVSM [677] for a vector space whose scalars come from an arbitrary field F. Contributed by Robert Beezer

T20 By applying Definition VS [659], prove that every field is also a vector space. (See the construction at the end of this section.) Contributed by Robert Beezer

Subsection SOL Solutions

C60 Contributed by Robert Beezer Statement [1841] Remember that every computation must be done with arithmetic in the field, reducing any intermediate number outside of $\{0, 1, 2, 3, 4\}$ to its remainder after division by 5.

The matrix inverse can be found with Theorem CINM [516] (and we discover along the way that A is nonsingular). The inverse is

$$A^{-1} = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 3 & 4 & 1 & 4 \\ 1 & 4 & 0 & 2 \\ 3 & 0 & 1 & 0 \end{bmatrix}$$

Then by an application of Theorem SNCM [543] the (unique) solution to the system will be

$$A^{-1}\mathbf{b} = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 3 & 4 & 1 & 4 \\ 1 & 4 & 0 & 2 \\ 3 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 2 \\ 0 \end{bmatrix} \qquad \qquad = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

Section T Trace

This section contributed by Andy Zimmer.

The matrix trace is a function that sends square matrices to scalars. In some ways it is reminiscent of the determinant. And like the determinant, it has many useful and surprising properties.

Definition T

Trace

Suppose A is a square matrix of size n. Then the **trace** of A, t(A), is the sum of the diagonal entries of A. Symbolically,

$$t\left(A\right) = \sum_{i=1}^{n} \left[A\right]_{ii}$$

(This definition contains Notation T.)

The next three proofs make for excellent practice. In some books they would be left as exercises for the reader as they are all "trivial" in the sense they do not rely on anything but the definition of the matrix trace.

 \wedge

Theorem TL Trace is Linear

Suppose A and B are square matrices of size n. Then t(A + B) = t(A) + t(B). Furthermore, if $\alpha \in \mathbb{C}$, then $t(\alpha A) = \alpha t(A)$.

Proof These properties are exactly those required for a linear transformation. To prove these results we just manipulate sums,

$$t (A + B) = \sum_{k=1}^{n} [A + B]_{ii}$$
 Definition T [1845]
$$= \sum_{i=1}^{n} [A]_{ii} + [B]_{ii}$$
 Definition MA [427]
$$= \sum_{i=1}^{n} [A]_{ii} + \sum_{i=1}^{n} [B]_{ii}$$
 Property CACN [1597]
$$= t (A) + t (B)$$
 Definition T [1845]

The second part is as straightforward as the first,

$$t(\alpha A) = \sum_{i=1}^{n} [\alpha A]_{ii}$$
$$= \sum_{i=1}^{n} \alpha [A]_{ii}$$

Definition T [1845]

Definition MSM [428]

$$= \alpha \sum_{i=1}^{n} [A]_{ii}$$
$$= \alpha t (A)$$

Property DCN [1597]

Definition T [1845]

[471]

[471]

Theorem TSRM Trace is Symmetric with Respect to Multiplication Suppose A and B are square matrices of size n. Then t(AB) = t(BA). Proof

$$t (AB) = \sum_{k=1}^{n} [AB]_{kk}$$
 Definition T [1845]
$$= \sum_{k=1}^{n} \sum_{\ell=1}^{n} [A]_{k\ell} [B]_{\ell k}$$
 Theorem EMP [471]
$$= \sum_{\ell=1}^{n} \sum_{k=1}^{n} [A]_{k\ell} [B]_{\ell k}$$
 Property CACN [1597]
$$= \sum_{\ell=1}^{n} \sum_{k=1}^{n} [B]_{\ell k} [A]_{k\ell}$$
 Property CMCN [1597]
$$= \sum_{\ell=1}^{n} [BA]_{\ell \ell}$$
 Theorem EMP [471]

 \square

= t (BA) Definition T [1845]

Theorem TIST Trace is Invariant Under Similarity Transformations

Suppose A and S are square matrices of size n and S is invertible. Then $t(S^{-1}AS) = t(A)$.

Proof Invariant means constant under some operation. In this case the operation is a similarity transformation. A lengthy exercise (but possibly a educational one) would be to prove this result without referencing Theorem TSRM [1847]. But here we will,

$t\left(S^{-1}AS\right) = t\left(\left(S^{-1}A\right)S\right)$	Theorem MMA $[478]$
$= t\left(S\left(S^{-1}A\right)\right)$	Theorem TSRM [1847]
$= t\left(\left(SS^{-1}\right)A\right)$	Theorem MMA [478]
$=t\left(A ight)$	Definition MI [506]

Now we could define the trace of a linear transformation as the trace of any matrix representation of the transformation. Would this definition be well-defined? That is, will two different representations of the same linear transformation always have the same trace? Why? (Think Theorem SCB [1377].) We will now prove one of the most interesting and surprising results about the trace.

Theorem TSE Trace is the Sum of the Eigenvalues

Suppose that A is a square matrix of size n with distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$. Then

$$t(A) = \sum_{i=1}^{k} \alpha_A(\lambda_i) \lambda_i$$

Proof It is amazing that the eigenvalues would have anything to do with the sum of the diagonal entries. Our proof will rely on double counting. We will demonstrate two different ways of counting the same thing therefore proving equality. Our object of interest is the coefficient of x^{n-1} in the characteristic polynomial of A (Definition CP [960]), which will be denoted α_{n-1} . From the proof of Theorem NEM [1015] we have,

$$p_A(x) = (-1)^n (x - \lambda_1)^{\alpha_A(\lambda_1)} (x - \lambda_2)^{\alpha_A(\lambda_2)} (x - \lambda_3)^{\alpha_A(\lambda_3)} \cdots (x - \lambda_k)^{\alpha_A(\lambda_k)}$$

First we want to prove that α_{n-1} is equal to $(-1)^{n+1} \sum_{i=1}^{k} \alpha_A(\lambda_i) \lambda_i$ and to do this we will use a straight forward counting argument. Induction can be used here as well (try it), but the intuitive approach is a much stronger technique. Let's imagine creating each term one by one from the extended product. How do we do this? From each $(x - \lambda_i)$ we pick either a x or a λ_i . But we are only interested in the terms that result in x to the power n - 1. As $\sum_{i=1}^{k} \alpha_A(\lambda_i) = n$, we have n factors of the form $(x - \lambda_i)$. Then to get terms with x^{n-1} we

 \square

need to pick x's in every $(x - \lambda_i)$, except one. Since we have n linear factors there are n ways to do this, namely each eigenvalue represented as many times as its algebraic multiplicity. Now we have to take into account the sign of each term. As we pick n - 1 x's and one λ_i (which has a negative sign in the linear factor) we get a factor of -1. Then we have to take into account the $(-1)^n$ in the characteristic polynomial. Thus α_{n-1} is the sum of these terms,

$$\alpha_{n-1} = (-1)^{n+1} \sum_{i=1}^{k} \alpha_A(\lambda_i) \lambda_i$$

Now we will now show that α_{n-1} is also equal to $(-1)^{n-1}t(A)$. For this we will proceed by induction on the size of A. If A is a 1×1 square matrix then $p_A(x) = \det(A - xI_n) = ([A]_{11} - x)$ and $(-1)^{1-1}t(A) = [A]_{11}$. With our base case in hand let's assume A is a square matrix of size n. By Definition CP [960]

$$p_A(x) = \det (A - xI_n)$$

= $[A - xI_n]_{11} \det ((A - xI_n)(1|1)) - [A - xI_n]_{12} \det ((A - xI_n)(1|2)) +$
 $[A - xI_n]_{13} \det ((A - xI_nn)(1|3)) - \dots + (-1)^{n+1} [A - xI_n]_{1n} \det ((A - xI_n)(1|n))$

First let's consider the maximum degree of $[A - xI_n]_{1i} \det ((A - xI_n)(1|i))$ when $i \neq 1$. For polynomials, the degree of f, denoted d(f), is the highest power of x in the expression f(x). A well known result of this definition is: if f(x) = g(x)h(x) then d(f) = d(g) + d(h) (can you prove this?). Now $[A - xI_n]_{1i}$ has degree zero when $i \neq 1$. Furthermore $(A - xI_n)(1|i)$ has n-1 rows, one of which has all of its entries of degree zero, since column *i* is removed. The other n-2 rows have one entry with degree one and the remainder of degree zero. Then by Exercise T.T30 [1853], the maximum degree of $[A - xI_n]_{1i} \det ((A - xI_n)(1|i))$ is n-2. So these terms will not affect the coefficient of x^{n-1} . Now we are free to focus all of our attention on the term $[A - xI_n]_{11} \det ((A - xI_n)(1|1))$. As A(1|1) is a $(n-1) \times (n-1)$ matrix the induction hypothesis tells us that det $((A - xI_n)(1|1))$ has a coefficient of $(-1)^{n-2}t(A(1|1))$ for x^{n-2} . We also note that the proof of Theorem NEM [1015] tells us that the leading coefficient of det $((A - xI_n)(1|1))$ is $(-1)^{n-1}$. Then,

$$[A - xI_n]_{11} \det \left((A - xI_n) \left(1 | 1 \right) \right) = \left([A]_{11} - x \right) \left((-1)^{n-1} x^{n-1} + (-1)^{n-2} t \left(A \left(1 | 1 \right) \right) x^{n-2} + \ldots \right)$$

Expanding the product shows α_{n-1} (the coefficient of x^{n-1}) to be

$$\begin{aligned} \alpha_{n-1} &= (-1)^{n-1} [A]_{11} + (-1)^{n-1} t \left(A \left(1 | 1 \right) \right) \\ &= (-1)^{n-1} [A]_{11} + (-1)^{n-1} \sum_{k=1}^{n-1} [A \left(1 | 1 \right)]_{kk} \qquad \text{Definition T [1845]} \\ &= (-1)^{n-1} \left([A]_{11} + \sum_{k=1}^{n-1} [A \left(1 | 1 \right)]_{kk} \right) \qquad \text{Property DCN [1597]} \\ &= (-1)^{n-1} \left([A]_{11} + \sum_{k=2}^{n} [A]_{kk} \right) \qquad \text{Definition SM [893]} \\ &= (-1)^{n-1} t \left(A \right) \qquad \text{Definition T [1845]} \end{aligned}$$

With two expressions for α_{n-1} , we have our result,

$$t (A) = (-1)^{n+1} (-1)^{n-1} t (A)$$

= $(-1)^{n+1} \alpha_{n-1}$
= $(-1)^{n+1} (-1)^{n+1} \sum_{i=1}^{k} \alpha_A (\lambda_i) \lambda_i$
= $\sum_{i=1}^{k} \alpha_A (\lambda_i) \lambda_i$

Subsection EXC Exercises

T10 Prove there are no square matrices A and B such that $AB - BA = I_n$. Contributed by Andy Zimmer

T12 Assume A is a square matrix of size n matrix. Prove $t(A) = t(A^t)$. Contributed by Andy Zimmer

T20 If $T_n = \{ M \in M_{nn} | t(M) = 0 \}$ then prove T_n is a subspace of M_{nn} and determine its dimension.

Contributed by Andy Zimmer

T30 Assume A is a $n \times n$ matrix with polynomial entries. Define md(A, i) to be the maximum degree of the entries in row i. Then $d(\det(A)) \leq md(A, 1) + md(A, 2) + \ldots + md(A, n)$. (Hint: If f(x) = h(x) + g(x), then $d(f) \leq \max\{d(h), d(g)\}$.) Contributed by Andy Zimmer Solution [1855]

T40 If A is a square matrix, the **matrix exponential** is defined as

$$e^A = \sum_{i=0}^{\infty} \frac{A^i}{i!}$$

Prove that det $(e^A) = e^{t(A)}$. (You might want to give some thought to the convergence of

the infinite sum as well.) Contributed by Andy Zimmer

Subsection SOL Solutions

T30 Contributed by Andy Zimmer Statement [1853] We will proceed by induction. If A is a square matrix of size 1, then clearly $d(\det(A)) \leq md(A, 1)$. Now assume A is a square matrix of size n then by Theorem DER [897],

$$\det (A) = (-1)^2 [A]_{1,1} \det (A(1|1)) + (-1)^3 [A]_{1,2} \det (A(1|2)) + (-1)^4 [A]_{1,3} \det (A(1|3)) + \dots + (-1)^{n+1} [A]_{1,n} \det (A(1|n))$$

Let's consider the degree of term j, $(-1)^{1+j} [A]_{1,j} \det (A(1|j))$. By definition of the function md, $d([A]_{1,j}) \leq md(A, j)$. We use our induction hypothesis to examine the other part of the product which tells us that

$$d\left(\det\left(A\left(1|j\right)\right)\right) \le md(A\left(1|j\right), 1) + md(A\left(1|j\right), 2) + \dots + md(A\left(1|j\right), n-1)$$

Furthermore by definition of A(1|j) (Definition SM [893]) row *i* of matrix A contains all the entries of the corresponding row in A(1|j) then,

 $md(A(1|j), 1) \le md(A, 1)$ $md(A(1|j), 2) \le md(A, 2)$

$$md(A(1|j), j-1) \le md(A, j-1)$$
$$md(A(1|j), j) \le md(A, j+1)$$
$$\vdots$$
$$md(A(1|j), n-1) \le md(A, n)$$

So,

$$d(\det(A(1|j))) \le md(A(1|j), 1) + md(A(1|j), 2) + \dots + md(A(1|j), n-1) \\ \le md(A, 1) + md(A, 2) + \dots + md(A, j-1) + md(A, j+1) + \dots + md(A, n-1)$$

Then using the property that if f(x) = g(x)h(x) then d(f) = d(g) + d(h),

$$d\left((-1)^{1+j} [A]_{1,j} \det (A(1|j))\right) = d\left([A]_{1,j}\right) + d\left(\det (A(1|j))\right)$$

$$\leq md(A, j) + md(A, 1) + md(A, 2) + \dots + md(A, j-1) + md(A, j-1) + md(A, j+1) + \dots + md(A, n)$$

$$= md(A, 1) + md(A, 2) + \dots + md(A, n)$$

As j is arbitrary the degree of all terms in the determinant are so bounded. Finally using the fact that if f(x) = g(x) + h(x) then $d(f) \le \max\{d(h), d(g)\}$ we have

$$d(\det(A)) \le md(A,1) + md(A,2) + \dots + md(A,n)$$

Section HP Hadamard Product

This section is contributed by Elizabeth Million.

You may have once thought that the natural definition for matrix multiplication would be entrywise multiplication, much in the same way that a young child might say, "I writed my name." The mistake is understandable, but it still makes us cringe. Unlike poor grammar, however, entrywise matrix multiplication has reason to be studied; it has nice properties in matrix analysis and additionally plays a role with relative gain arrays in chemical engineering, covariance matrices in probability and serves as an inertia preserver for Hermitian matrices in physics. Here we will only explore the properties of the Hadamard product in matrix analysis.

Definition HP Hadamard Product

Let A and B be $m \times n$ matrices. The **Hadamard Product** of A and B is defined by $[A \circ B]_{ij} = [A]_{ij} [B]_{ij}$ for all $1 \le i \le m, 1 \le j \le n$. (This definition contains Notation HP.)

As we can see, the Hadamard product is simply "entrywise multiplication". Because of this, the Hadamard product inherits the same benefits (and restrictions) of multiplication in \mathbb{C} . Note also that both A and B need to be the same size, but not necessarily square. To

avoid confusion, juxtaposition of matrices will imply the "usual" matrix multiplication, and we will use "o" for the Hadamard product.

Example HP Hadamard Product

Consider

$$A = \begin{bmatrix} 1 & 0 & 6 \\ 3 & \pi & 5 \end{bmatrix} \qquad \qquad B = \begin{bmatrix} 3 & 13 & i \\ \frac{1}{3} & 2 & 4 \end{bmatrix}$$

Then

$$A \circ B = \begin{bmatrix} (1)(3) & (0)(13) & (6)(i) \\ 3(\frac{1}{3}) & (\pi)(2) & (5)(4) \end{bmatrix}$$
$$= \begin{bmatrix} 3 & 0 & 6i \\ 1 & 2\pi & 20 \end{bmatrix}.$$

Now we will	explore some	basics pr	operties of	the I	Hadamard	Product.
-------------	--------------	-----------	-------------	-------	----------	----------

Theorem HPC Hadamard Product is Commutative

If A and B are $m \times n$ matrices then $A \circ B = B \circ A$.

Proof The proof follows directly from the fact that multiplication in \mathbb{C} is commutative. Let A and B be $m \times n$ matrices. Then

$$[A \circ B]_{ij} = [A]_{ij} [B]_{ij}$$
 Definition HP [1857]

 \boxtimes

 \square

$$= [B]_{ij} [A]_{ij}$$
Property CMCN [1597]
$$= [B \circ A]_{ij}$$
Definition HP [1857]

Definition HID Hadamard Identity The **Hadamard identity** is the $m \times n$ matrix J_{mn} defined by $[J_{mn}]_{ij} = 1$ for all $1 \le i \le m$, $1 \le j \le n$. (This definition contains Notation HID.) \triangle

Theorem HPHID Hadamard Product with the Hadamard Identity Suppose A is an $m \times n$ matrix. Then $A \circ J_{mn} = J_{mn} \circ A = A$.

Proof

$$[A \circ J_{mn}]_{ij} = [J_{mn} \circ A]_{ij}$$

$$= [J_{mn}]_{ij} [A]_{ij}$$

$$= (1) [A]_{ij}$$

$$= [A]_{ij}$$
Theorem HPC [1858]
Definition HP [1857]
Definition HID [1859]
Property OCN [1597]

Definition HI Hadamard Inverse

Let A be an $m \times n$ matrix and suppose $[A]_{ij} \neq 0$ for all $1 \leq i \leq m, 1 \leq j \leq n$. Then the **Hadamard Inverse**, \widehat{A} , is given by $[\widehat{A}]_{ij} = ([A]_{ij})^{-1}$ for all $1 \leq i \leq m, 1 \leq j \leq n$. (This definition contains Notation HI.) \bigtriangleup

Theorem HPHI

Hadamard Product with Hadamard Inverses

Let A be an $m \times n$ matrix such that $[A]_{ij} \neq 0$ for all $1 \leq i \leq m, 1 \leq j \leq n$. Then $A \circ \widehat{A} = \widehat{A} \circ A = J_{mn}$.

Proof

$$\begin{bmatrix} A \circ \widehat{A} \end{bmatrix}_{ij} = \begin{bmatrix} \widehat{A} \circ A \end{bmatrix}_{ij}$$
$$= \begin{bmatrix} \widehat{A} \end{bmatrix}_{ij} \begin{bmatrix} A \end{bmatrix}_{ij}$$
$$= (\begin{bmatrix} A \end{bmatrix}_{ij})^{-1} \begin{bmatrix} A \end{bmatrix}_{ij}$$
$$= 1$$
$$= \begin{bmatrix} J_{mn} \end{bmatrix}_{ij}$$

Theorem HPC [1858]

Definition HP [1857]

Definition HI [1860], $[A]_{ij} \neq 0$ Property MICN [1598] Definition HID [1859]

Since matrices have a different inverse and identity under the Hadamard product, we have used special notation to distinguish them from what we have been using with "normal" matrix multiplication. That is, compare "usual" matrix inverse, A^{-1} , with the Hadamard inverse \hat{A} , and the "usual" matrix identity, I_n , with the Hadamard identity, J_{mn} . The Hadamard identity matrix and the Hadamard inverse are both more limiting than helpful, so we will not explore their use further. One last fun fact for those of you who may be familiar with group theory: the set of $m \times n$ matrices with nonzero entries form an abelian (commutative) group under the Hadamard product (prove this!).

Theorem HPDAA Hadamard Product Distributes Across Addition

Suppose A, B and C are $m \times n$ matrices. Then $C \circ (A + B) = C \circ A + C \circ B$.

Proof

$$\begin{split} [C \circ (A+B)]_{ij} &= [C]_{ij} [A+B]_{ij} & \text{Definition HP [1857]} \\ &= [C]_{ij} ([A]_{ij} + [B]_{ij}) & \text{Definition MA [427]} \\ &= [C]_{ij} [A]_{ij} + [C]_{ij} [B]_{ij} & \text{Property DCN [1597]} \\ &= [C \circ A]_{ij} + [C \circ B]_{ij} & \text{Definition HP [1857]} \\ &= [C \circ A + C \circ B]_{ij} & \text{Definition MA [427]} \end{split}$$

Theorem HPSMM Hadamard Product and Scalar Matrix Multiplication Suppose $\alpha \in \mathbb{C}$, and A and B are $m \times n$ matrices. Then $\alpha(A \circ B) = (\alpha A) \circ B = A \circ (\alpha B)$. \Box

Proof

$$[\alpha A \circ B]_{ij} = \alpha [A \circ B]_{ij}$$

$$= \alpha [A]_{ij} [B]_{ij}$$

$$= [\alpha A]_{ij} [B]_{ij}$$

$$= [(\alpha A) \circ B]_{ij}$$

$$= \alpha [A]_{ij} [B]_{ij}$$

$$= [A]_{ij} \alpha [B]_{ij}$$

$$= [A]_{ij} [\alpha B]_{ij}$$

$$= [A \circ (\alpha B)]_{ij}$$

Definition MSM [428] Definition HP [1857] Definition MSM [428] Definition HP [1857] Definition MSM [428] Property CMCN [1597] Definition MSM [428] Definition HP [1857]

With equality of each entry of the matrices being equal we know by Definition ME [427] that the two matrices are equal.

Subsection DMHP Diagonal Matrices and the Hadamard Product

We can relate the Hadamard product with matrix multiplication by considering diagonal matrices, since $A \circ B = AB$ if and only if both A and B are diagonal. (Can you prove this?) For example, a simple calculation reveals that the Hadamard product relates the diagonal values of a diagonalizable matrix A with its eigenvalues:

Theorem DMHP

Diagonalizable Matrices and the Hadamard Product

Let A be a diagonalizable matrix of size n with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$. Let D be a diagonal matrix from the diagonalization of A, $A = SDS^{-1}$, and **d** be a vector such that $[D]_{ii} = [\mathbf{d}]_i = \lambda_i$ for all $1 \leq i \leq n$. Then

$$[A]_{ii} = \left[S \circ (S^{-1})^t \mathbf{d}\right]_i \qquad \qquad \text{for all } 1 \le i \le n.$$

That is,

$$\begin{bmatrix} [A]_{11} \\ [A]_{22} \\ [A]_{33} \\ \vdots \\ [A]_{nn} \end{bmatrix} = S \circ (S^{-1})^t \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_n \end{bmatrix}$$

Proof

$$\begin{bmatrix} S \circ (S^{-1})^{t} \mathbf{d} \end{bmatrix}_{i} = \sum_{k=1}^{n} \begin{bmatrix} S \circ (S^{-1})^{t} \end{bmatrix}_{ik} [\mathbf{d}]_{k} \qquad \text{Definition MVP [461]} \\ = \sum_{k=1}^{n} \begin{bmatrix} S \circ (S^{-1})^{t} \end{bmatrix}_{ik} \lambda_{k} \qquad \text{Definition of } \mathbf{d} \\ = \sum_{k=1}^{n} \begin{bmatrix} S \end{bmatrix}_{ik} \begin{bmatrix} (S^{-1})^{t} \end{bmatrix}_{ik} \lambda_{k} \qquad \text{Definition HP [1857]} \\ = \sum_{k=1}^{n} \begin{bmatrix} S \end{bmatrix}_{ik} \begin{bmatrix} S^{-1} \end{bmatrix}_{ki} \lambda_{k} \qquad \text{Definition TM [433]} \\ = \sum_{k=1}^{n} \begin{bmatrix} S \end{bmatrix}_{ik} \lambda_{k} \begin{bmatrix} S^{-1} \end{bmatrix}_{ki} \qquad \text{Property CMCN [1597]} \\ = \sum_{k=1}^{n} \begin{bmatrix} S \end{bmatrix}_{ik} \begin{bmatrix} D \end{bmatrix}_{kk} \begin{bmatrix} S^{-1} \end{bmatrix}_{ki} \qquad \text{Definition of } D \\ = \sum_{j=1}^{n} \sum_{k=1}^{n} \begin{bmatrix} S \end{bmatrix}_{ik} \begin{bmatrix} D \end{bmatrix}_{kj} \begin{bmatrix} S^{-1} \end{bmatrix}_{ji} \qquad \begin{bmatrix} D \end{bmatrix}_{kj} = 0 \text{ for all } k \neq j \end{bmatrix}$$

$$= \sum_{j=1}^{n} [SD]_{ij} [S^{-1}]_{ji}$$
 Theorem EMP [471]
$$= [SDS^{-1}]_{ii}$$
 Theorem EMP [471]
$$= [A]_{ii}$$
 Definition ME [427]

We obtain a similar result when we look at the singular value decomposition of square matrices (see exercises).

Theorem DMMP

Diagonal Matrices and Matrix Products

Suppose A, B are $m \times n$ matrices, and D and E are diagonal matrices of size m and n, respectively. Then,

$$D(A \circ B)E = (DAE) \circ B = (DA) \circ (BE)$$

Proof

$$[D(A \circ B)E]_{ij} = \sum_{k=1}^{m} [D]_{ik} [(A \circ B)E]_{kj}$$
 Theorem EMP [471]
$$= \sum_{k=1}^{m} \sum_{l=1}^{n} [D]_{ik} [A \circ B]_{kl} [E]_{lj}$$
 Theorem EMP [471]

 \square

$$\begin{split} &= \sum_{k=1}^{m} \sum_{l=1}^{n} [D]_{ik} [A]_{kl} [B]_{kl} [E]_{lj} & \text{Definition HP [1857]} \\ &= \sum_{k=1}^{m} [D]_{ik} [A]_{kj} [B]_{kj} [E]_{jj} & [E]_{lj} = 0 \text{ for all } l \neq j \\ &= [D]_{ii} [A]_{ij} [B]_{ij} [E]_{jj} & [D]_{ik} = 0 \text{ for all } i \neq k \\ &= [D]_{ii} [A]_{ij} [E]_{jj} [B]_{ij} & Property CMCN [1597] \\ &= [D]_{ii} (\sum_{l=1}^{n} [A]_{il} [E]_{lj}) [B]_{ij} & [E]_{lj} = 0 \text{ for all } l \neq j \\ &= [D]_{ii} [AE]_{ij} [B]_{ij} & Theorem EMP [471] \\ &= (\sum_{k=1}^{m} [D]_{ik} [AE]_{kj}) [B]_{ij} & [D]_{ik} = 0 \text{ for all } i \neq k \\ &= [DAE]_{ij} [B]_{ij} & Theorem EMP [471] \\ &= [(DAE) \circ B]_{ij} & Theorem EMP [471] \\ &= [(DAE) \circ B]_{ij} & Definition HP [1857] \end{split}$$

Also,

$$[(DAE) \circ B]_{ij} = [DAE]_{ij} [B]_{ij}$$
 Definition HP [1857]

$$= (\sum_{k=1}^{n} [DA]_{ik} [E]_{kj}) [B]_{ij}$$
Theorem EMP [471]

$$= [DA]_{ij} [E]_{jj} [B]_{ij}$$
[E]_{kj} = 0 for all $k \neq j$

$$= [DA]_{ij} [B]_{ij} [E]_{jj}$$
Property CMCN [1597]

$$= [DA]_{ij} (\sum_{k=1}^{n} [B]_{ik} [E]_{kj})$$
[E]_{kj} = 0 for all $k \neq j$

$$= [DA]_{ij} [BE]_{ij}$$
Theorem EMP [471]

$$= [(DA) \circ (BE)]_{ij}$$
Definition HP [1857]

Subsection EXC Exercises

T10 Prove that $A \circ B = AB$ if and only if both A and B are diagonal matrices. Contributed by Elizabeth Million

T20 Suppose A, B are $m \times n$ matrices, and D and E are diagonal matrices of size m and n, respectively. Prove both parts of the following equality hold:

$$D(A \circ B)E = (AE) \circ (DB) = A \circ (DBE)$$

Contributed by Elizabeth Million

T30 Let A be a square matrix of size n with singular values $\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n$. Let D be a diagonal matrix from the singular value decomposition of A, $A = UDV^*$ (Theorem SVD [1924]). Define the vector **d** by $[\mathbf{d}]_i = [D]_{ii} = \sigma_i, 1 \le i \le n$. Prove the following equality,

$$\left[A\right]_{ii} = \left[\left(U \circ \overline{V}\right)\mathbf{d}\right]_i$$

Contributed by Elizabeth Million

T40 Suppose A, B and C are $m \times n$ matrices. Prove that for all $1 \le i \le m$, $\left[(A \circ B)C^t \right]_{ii} = \left[(A \circ C)B^t \right]_{ii}$ Contributed by Elizabeth Million

T50 Define the diagonal matrix D of size n with entries from a vector $\mathbf{x} \in \mathbb{C}^n$ by

$$[D]_{ij} = \begin{cases} [x]_i & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, suppose A, B are $m \times n$ matrices. Prove that $[ADB^t]_{ii} = [(A \circ B)\mathbf{x}]_i$ for all $1 \leq i \leq m$. Contributed by Elizabeth Million

Section VM Vandermonde Matrix

This Section is a Draft, Subject to Changes

Alexandre-Théophile Vandermonde was a French mathematician in the 1700's who was among the first to write about basic properties of the determinant (such as the effect of swapping two rows). However, the determinant that bears his name (Theorem DVM [1871]) does not appear in any of his four published mathematical papers.

Definition VM Vandermonde Matrix

An square matrix of size n, A, is a **Vandermonde matrix** if there are scalars, $x_1, x_2, x_3, \ldots, x_n$ such that $[A]_{ij} = x_i^{j-1}, 1 \le i \le n, 1 \le j \le n$.

Example VM4 Vandermonde matrix of size 4

$$A = \begin{bmatrix} 1 & 2 & 4 & 8 \\ 1 & -3 & 9 & -27 \\ 1 & 1 & 1 & 1 \\ 1 & 4 & 16 & 64 \end{bmatrix}$$

is a Vandermonde matrix since it meets the definition with $x_1 = 2, x_2 = -3, x_3 = 1, x_4 = 4$.

Vandermonde matrices are not very interesting as numerical matrices, but instead appear more often in proofs and applications where the scalars x_i are carried as symbols. Two such applications are in the sections on secret-sharing (Section SAS [1949]) and curve-fitting (Section CF [1940]). Principally, we would like to know when Vandermonde matrices are nonsingular, and the most convenient way to check this is by determining when the determinant is nonzero (Theorem SMZD [935]). As a bonus, the determinant of a Vandermonde matrix has an especially pleasing formula.

Theorem DVM

Determinant of a Vandermonde Matrix

Suppose that A is a Vandermonde matrix of size n built with the scalars $x_1, x_2, x_3, \ldots, x_n$. Then

$$\det(A) = \prod_{1 \le i < j \le n} (x_j - x_i)$$

Proof The proof is by induction (Technique I [1630]) on n, the size of the matrix. An empty product for a 1×1 matrix might make a good base case, but we'll start at n = 2 instead. For a 2×2 Vandermonde matrix, we have

$$\det (A) = \begin{vmatrix} 1 & x_1 \\ 1 & x_2 \end{vmatrix} = x_2 - x_1 = \prod_{1 \le i < j \le 2} (x_j - x_i)$$

 \square

For the induction step we will perform row operations on A to obtain the determinant of A as multiple of the determinant of an $(n-1) \times (n-1)$ Vandermonde matrix. the notation in this theorem tens to obscure your intuition about the changes effected by various row and column manipulations. Construct a 4×4 Vandermonde matrix with four symbols as the scalars $(x_1, x_2, x_2, x_4, \text{ or perhaps } a, b, c, d)$ and play along with the example as you study the proof.

First we convert most of the first column to zeros. Subtract row n from each of the other n-1 rows to form a matrix B. By Theorem DRCMA [924], B has the same determinant as A. The entries of B, in the first n-1 rows, i.e. for $1 \le i \le n-1$, $1 \le j \le n-1$, are

$$[B]_{ij} = x_i^{j-1} - x_n^{j-1} = (x_i - x_n) \sum_{k=0}^{j-2} x_i^{j-2-k} x_n^k$$

As the elements of row $i, 1 \leq i \leq n-1$, have the common factor $(x_i - x_n)$, we form the new matrix C that differs from B by the removal of this factor from each of the first n-1 rows. This will change the determinant, as we will track carefully in a moment. We also have a first column with zeros in each location, except row n, so we can use it for a column expansion computation of the determinant. We now know,

$$det (A) = det (B)$$

$$= (x_1 - x_n)(x_2 - x_n) \cdots (x_{n-1} - x_n) det (C)$$
Theorem DRCMA [924]
$$= (x_1 - x_n)(x_2 - x_n) \cdots (x_{n-1} - x_n)(1)(-1)^{n+1} det (C (n-1|1))$$
Theorem DEC [901]
$$= (x_1 - x_n)(x_2 - x_n) \cdots (x_{n-1} - x_n)(-1)^{n-1} det (C (n-1|1))$$

$$= (x_n - x_1)(x_n - x_2) \cdots (x_n - x_{n-1}) \det (C(n-1|1))$$

For convenience, denote D = C(n-1|1). Entries of this matrix are similar to those of B, but the factors used to build C are gone, and since the first column is gone, there is a slight re-indexing relative to the columns. For $1 \le i \le n-1$, $1 \le j \le n-1$,

$$[D]_{ij} = \sum_{k=0}^{j-1} x_i^{j-1-k} x_n^k$$

We will perform many column operations on the matrix D, always of the type where we multiply a column by a scalar and add the result to another column. As such, Theorem DRCM [922] insures that the determinant will remain constant. We will work column by column, left to right, to convert D into a Vandermonde matrix with scalars $x_1, x_2, x_3, \ldots, x_{n-1}$. More precisely, we will build a sequence of matrices $D = D_1, D_2, \ldots, D_{n-1}$, where each obtainable from the previous by a sequence of determinant-preserving column operations and the first ℓ columns of D_{ℓ} are the first ℓ columns of a Vandermonde matrix with scalars $x_1, x_2, x_3, \ldots, x_{n-1}$. We could establish this claim by induction (Technique I [1630]) on ℓ if we were to expand the claim to specify the exact values of the final $n - 1 - \ell$ columns as well. Since the claim is that matrices with certain properties exist, we will instead establish the claim by constructing the desired matrices one-by-one procedurally. The extension to an inductive proof should be clear, but not especially illuminating.

Set $D_1 = D$ to begin, and note that the entries of the first column of D_1 are, for

 $1 \le i \le n-1,$

$$[D_1]_{i1} = \sum_{k=0}^{1-1} x_i^{1-1-k} x_n^k = 1 = x_i^{1-1}$$

So the first column of D_1 has the properties we desire. We will use this column of all 1's to remove the highest power of x_n from each of the remaining columns and so build D_2 . Precisely, perform the n-2 column operations where column 1 is multiplied by x_n^{j-1} and subtracted from column j, for $2 \le j \le n-1$. Call the result D_2 , and examine its entries in columns 2 through n-1. For $1 \le i \le n-1$, $2 \le j \le n-1$,

$$D_{2}]_{ij} = -x_{n}^{j-1} [D_{1}]_{i1} + [D_{1}]_{ij}$$

$$= -x_{n}^{j-1} (1) + \sum_{k=0}^{j-1} x_{i}^{j-1-k} x_{n}^{k}$$

$$= -x_{n}^{j-1} + x_{i}^{j-1-(j-1)} x_{n}^{j-1} + \sum_{k=0}^{j-2} x_{i}^{j-1-k} x_{n}^{k}$$

$$= \sum_{k=0}^{j-2} x_{i}^{j-1-k} x_{n}^{k}$$

In particular, we examine column 2 of D_2 . For $1 \le i \le n-1$,

$$[D_2]_{i2} = \sum_{k=0}^{2-2} x_i^{2-1-k} x_n^k = x_i^1 = x_i^{2-1}$$

Now, form D_3 . Perform the n-3 column operations where column 2 of D_2 is multiplied by x_n^{j-2} and subtracted from column j, for $3 \le j \le n-1$. The result is D_3 , whose entries we now compute. For $1 \le i \le n-1$,

$$D_{3}]_{ij} = -x_{n}^{j-2} [D_{2}]_{i2} + [D_{2}]_{ij}$$

$$= -x_{n}^{j-2} x_{i}^{1} + \sum_{k=0}^{j-2} x_{i}^{j-1-k} x_{n}^{k}$$

$$= -x_{n}^{j-2} x_{i}^{1} + x_{i}^{j-1-(j-2)} x_{n}^{j-2} + \sum_{k=0}^{j-3} x_{i}^{j-1-k} x_{n}^{k}$$

$$= \sum_{k=0}^{j-3} x_{i}^{j-1-k} x_{n}^{k}$$

Specifically, we examine column 3 of D_3 . For $1 \le i \le n-1$,

$$[D_3]_{i3} = \sum_{k=0}^{3-3} x_i^{3-1-k} x_n^k = x_i^2 = x_i^{3-1}$$

We could continue this procedure n - 4 more times, eventually totaling $\frac{1}{2}(n^2 - 3n + 2)$ column operations, and arriving at D_{n-1} , the Vandermonde matrix of size n - 1 built from the scalars $x_1, x_2, x_3, \ldots, x_{n-1}$. Informally, we chop off the last term of every sum, until a single term is left in a column, and it is of the right form for the Vandermonde matrix.

This desired column is then used in the next iteration to chop off some more final terms for columns to the right. Now we can apply our induction hypothesis to the determinant of D_{n-1} and arrive at an expression for det A,

$$\det (A) = \det (C)$$

= $\prod_{k=1}^{n-1} (x_n - x_k) \det (D)$
= $\prod_{k=1}^{n-1} (x_n - x_k) \det (D_{n-1})$
= $\prod_{k=1}^{n-1} (x_n - x_k) \prod_{1 \le i < j \le n-1} (x_j - x_i)$
= $\prod_{1 \le i < j \le n} (x_j - x_i)$

which is the desired result.

Before we had Theorem DVM [1871] we could see that if two of the scalar values were equal, then the Vandermonde matrix would have two equal rows and hence be singular (Theorem DERC [923], Theorem SMZD [935]). But with this expression for the determinant, we can establish the converse.

Theorem NVM

Nonsingular Vandermonde Matrix

A Vandermonde matrix of size n with scalars $x_1, x_2, x_3, \ldots, x_n$ is nonsingular if and only if the scalars are all different.

Proof Let A denote the Vandermonde matrix with scalars $x_1, x_2, x_3, \ldots, x_n$. By Theorem SMZD [935], A is nonsingular if and only if the determinant of A is nonzero. The determinant is given by Theorem DVM [1871], and this product is nonzero if and only if each term of the product is nonzero. This condition translates to $x_i - x_j \neq 0$ whenever $i \neq j$. In other words, the matrix is nonsingular if and only if the scalars are all different.

Section PSM Positive Semi-definite Matrices

This Section is a Draft, Subject to Changes Needs Numerical Examples Inner product is "reversed" from prior material, see changelog Explanation

Positive semi-definite matrices (and their cousins, positive definite matrices) are square matrices which in many ways behave like non-negative (respectively, positive) real numbers. Results given here are employed in the decompositions of Section SVD [1914], Section SR [1926] and Section PD [850].

Subsection PSM Positive Semi-Definite Matrices

Definition PSM Positive Semi-Definite Matrix

A square matrix A of size n is **positive semi-definite** if A is Hermitian and for all $\mathbf{x} \in \mathbb{C}^n$, $\langle A\mathbf{x}, \mathbf{x} \rangle \geq 0$.

For a definition of **positive definite** replace the inequality in the definition with a strict inequality, and exclude the zero vector from the vectors \mathbf{x} required to meet the condition. Similar variations allow definitions of **negative definite** and **negative semi-definite**. Our first theorem in this section gives us an easy way to build positive semi-definite matrices.

Theorem CPSM

Creating Positive Semi-Definite Matrices

Suppose that A is any $m \times n$ matrix. Then the matrices A^*A and AA^* are positive semidefinite matrices.

Proof We will give the proof for the first matrix, the proof for the second is entirely similar. First we check that A^*A is Hermitian,

$(A^*A)^* = A^* (A^*)^*$	Theorem MMAD [483]
$= A^*A$	Theorem AA [443]

so by Definition HM [486], the matrix A^*A is Hermitian. Second, for any $\mathbf{x} \in \mathbb{C}^n$,

$\langle A^*A\mathbf{x}, \mathbf{x} \rangle = \langle A\mathbf{x}, (A^*)^* \mathbf{x} \rangle$	Theorem AIP $[485]$
$=\langle A\mathbf{x},A\mathbf{x} angle$	Theorem AA [443]
≥ 0	Theorem PIP [406]

which is the second criteria in the definition of a positive semi-definite matrix (Definition PSM [1879]).

A statement very similar to the converse of this theorem is also true. Any positive semidefinite matrix can be realized as the product of a square matrix, B, with its adjoint, B^* . (See Exercise PSM.T20 [1884] after studying this entire section.) The matrices A^*A and AA^* will be important later when we define singular values (Section SVD [1914]).

Positive semi-definite matrices can also be characterized by their eigenvalues, without any mention of inner products. This next result further reinforces the notion that positive semi-definite matrices behave like non-negative real numbers.

Theorem EPSM

Eigenvalues of Positive Semi-definite Matrices

Suppose that A is a Hermitian matrix. Then A is positive semi-definite matrix if and only if whenever λ is an eigenvalue of A, then $\lambda \ge 0$.

Proof Notice first that since we are considering only Hermitian matrices in this theorem, it is always possible to compare eigenvalues with the real number zero, since eigenvalues of Hermitian matrices are all real numbers (Theorem HMRE [1020]). Let n denote the size of A.

(⇒) Let $\mathbf{x} \neq 0$ be an eigenvector of A for λ . Then by Theorem PIP [406] we know $\langle \mathbf{x}, \mathbf{x} \rangle \neq 0$. So

$$\lambda = \frac{1}{\langle \mathbf{x}, \, \mathbf{x} \rangle} \lambda \langle \mathbf{x}, \, \mathbf{x} \rangle$$
$$= \frac{1}{\langle \mathbf{x}, \, \mathbf{x} \rangle} \langle \lambda \mathbf{x}, \, \mathbf{x} \rangle$$
$$= \frac{1}{\langle \mathbf{x}, \, \mathbf{x} \rangle} \langle A \mathbf{x}, \, \mathbf{x} \rangle$$

Property MICN [1598]

Theorem IPSM [401]

Definition EEM [945]

By Theorem PIP [406], $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ and by Definition PSM [1879] we have $\langle A\mathbf{x}, \mathbf{x} \rangle \ge 0$. With λ expressed as the product of these two quantities, we have $\lambda \ge 0$.

(\Leftarrow) Suppose now that $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ are the (not necessarily distinct) eigenvalues of the Hermitian matrix A, each of which is non-negative. Let $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$ be a set of associated eigenvectors for these eigenvalues. Since a Hermitian matrix is normal (Definition HM [486], Definition NM [173]), Theorem OBNM [1434] allows us to choose this set of eigenvectors to also be an orthonormal basis of \mathbb{C}^n . Choose any $\mathbf{x} \in \mathbb{C}^n$ and let $a_1, a_2, a_3, \ldots, a_n$ be the scalars guaranteed by the spanning property of the basis B such that

$$\mathbf{x} = a_1 \mathbf{x}_1 + a_2 \mathbf{x}_2 + a_3 \mathbf{x}_3 + \dots + a_n \mathbf{x}_n = \sum_{i=1}^n a_i \mathbf{x}_i$$

Since we have presumed A is Hermitian, we need only check the other defining property,

$$\langle A\mathbf{x}, \mathbf{x} \rangle = \left\langle A \sum_{i=1}^{n} a_i \mathbf{x}_i, \sum_{j=1}^{n} a_j \mathbf{x}_j \right\rangle$$
 Definition TSVS [746]

$$= \left\langle \sum_{i=1}^{n} A a_i \mathbf{x}_i, \sum_{j=1}^{n} a_j \mathbf{x}_j \right\rangle$$
 Theorem MMDAA [476]

$$= \left\langle \sum_{i=1}^{n} a_i A \mathbf{x}_i, \sum_{j=1}^{n} a_j \mathbf{x}_j \right\rangle$$
 Definition EEM [945]

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle a_i \lambda_i \mathbf{x}_i, a_j \mathbf{x}_j \rangle$$
 Theorem IPVA [400]

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \lambda_i \overline{a_j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$
 Theorem IPSM [401]

$$= \sum_{i=1}^{n} a_i \lambda_i \overline{a_i} \langle \mathbf{x}_i, \mathbf{x}_i \rangle + \sum_{i=1}^{n} \sum_{\substack{j=1\\ j \neq i}}^{n} a_i \lambda_i \overline{a_j} \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$
 Property CACN [1597]

$$= \sum_{i=1}^{n} a_i \lambda_i \overline{a_i}(1) + \sum_{i=1}^{n} \sum_{\substack{j=1\\ j \neq i}}^{n} a_i \lambda_i \overline{a_j}(0)$$
 Definition ONS [418]
$$= \sum_{i=1}^{n} a_i \lambda_i \overline{a_i}$$

$$= \sum_{i=1}^{n} \lambda_i |a_i|^2$$
 Definition MCN [1601]

With non-negative values for each eigenvalue λ_i , $1 \leq i \leq n$, and each modulus squared, it should be clear that this sum is non-negative. Which is exactly what is required by Definition PSM [1879] to establish that A is positive semi-definite.

As positive semi-definite matrices are defined to be Hermitian, they are then normal and subject to orthonormal diagonalization (Theorem OD [1429]). Now consider the interpretation of orthonormal diagonalization as a rotation to principal axes, a stretch by a diagonal matrix and a rotation back (Subsection OD.OD [1428]). For a positive semi-definite matrix, the diagonal matrix has diagonal entries that are the non-negative eigenvalues of the original positive semi-definite matrix. So the "stretching" along each axis is never a reflection.

Subsection EXC Exercises

T20 Suppose that A is a positive semi-definite matrix of size n. Prove that there is a square matrix B of size n such that $A = BB^*$. Contributed by Robert Beezer

Chapter MD Matrix Decompositions

INNER PRODUCT IS "REVERSED" FROM PRIOR MATERIAL, SEE CHANGELOG EXPLANATION

This chapter is about breaking up a matrix A into pieces that somehow combine to recreate A. Usually the pieces are again matrices, and usually they are then combined via matrix multiplication (Definition MM [468]). In some cases, the decomposition will be valid for any matrix, but often we might need extra conditions on A, such as being square (Definition SQM [173]), nonsingular (Definition NM [173]) or diagonalizable (Definition DZM [1037]) before we can guarantee the decomposition. If you are comfortable with topics like decomposing a solution vector into linear combinations (Subsection LC.VFSS [232]) or decomposing vector spaces into direct sums (Subsection PD.DS [863]), then we will be doing

similar things in this chapter. If not, review these ideas and take another look at Technique DC [1628] on decompositions.

We have studied one matrix decomposition already, so we will review that here in this introduction, both as a way of previewing the topic in a familiar setting, but also since it does not deserve another section all of its own.

A diagonalizable matrix (Definition DZM [1037]) is defined to be a square matrix A such that there is an invertible matrix S and a diagonal matrix D where $S^{-1}AS = D$. We can re-write this as $A = SDS^{-1}$. Here we have a decomposition of A into three matrices, S, Dand S^{-1} , which recombine through matrix multiplication to recreate A. We also know that the diagonal entries of D are the eigenvalues of A. We cannot form this decomposition for just any matrix — A must be square and we know from Theorem DC [1039] that a matrix of size n is diagonalizable if and only if there is a basis for \mathbb{C}^n composed entirely of eigenvectors of A, or by Theorem DMFE [1044] we know that A is diagonalizable if and only if each eigenvalue of A has a geometric multiplicity equal to its algebraic multiplicity. Some authors prefer to call this an **eigen decomposition** of A rather than a **matrix diagonalization**.

Another decomposition, which is similar in flavor to matrix diagonalization, is orthonormal diagonalization (Theorem OD [1429]). Here we require the matrix A to be normal and we get the decomposition $A = UDU^*$, where D is a diagonal matrix with the eigenvalues of A on the diagonal, and U is unitary. The hypothesis that A is normal guarantees the decomposition and we get the extra information that U is unitary.

Each section of this chapter features a different matrix decomposition, with the exception of Section PSM [1878], which presents background information on positive semi-definite matrices required for singular value decompositions, square roots and polar decompositions.

Section ROD Rank One Decomposition

This Section is a Draft, Subject to Changes

Our first decomposition applies only to diagonalizable (Definition DZM [1037]) matrices, and yields a decomposition into a sum of very simple matrices.

Theorem ROD

Rank One Decomposition

Suppose that A is a diagonalizable matrix of size n and rank r. Then there are r square matrices $A_1, A_2, A_3, \ldots, A_r$, each of size n and rank 1 such that

$$A = A_1 + A_2 + A_3 + \dots + A_r$$

Furthermore, if $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_r$ are the nonzero eigenvalues of A, then there are two sets of r linearly independent vectors from \mathbb{C}^n ,

$$X = \{ \mathbf{x}_1, \, \mathbf{x}_2, \, \mathbf{x}_3, \, \dots, \, \mathbf{x}_r \} \qquad \qquad Y = \{ \mathbf{y}_1, \, \mathbf{y}_2, \, \mathbf{y}_3, \, \dots, \, \mathbf{y}_r \}$$

such that $A_k = \lambda_k \mathbf{x}_k \mathbf{y}_k^t$, $1 \le k \le r$.

Proof The proof is constructive. Generally, we will diagonalize A, creating a nonsingular matrix S and a diagonal matrix D. Then we split up the diagonal matrix into a sum of matrices with a single nonzero entry (on the diagonal). This fundamentally creates the decomposition in the statement of the theorem, the remainder is just bookkeeping. The vectors in X and Y will result from the columns of S and the rows of S^{-1} .

Let $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$ be the eigenvalues of A (repeated according to their algebraic multiplicity). If A has rank r, then dim $(\mathcal{N}(A)) = n - r$ (Theorem RPNC [831]). The null space of A is the eigenspace of the eigenvalue $\lambda = 0$ (Theorem EMNS [965]), so it follows that the algebraic multiplicity of $\lambda = 0$ is n - r, $\alpha_A(0) = n - r$. Presume that the complete list of eigenvalues is ordered so that $\lambda_k = 0$ for $r + 1 \le k \le n$.

Since A is hypothesized to be diagonalizable, there exists a diagonal matrix D and an invertible matrix S, such that $D = S^{-1}AS$. We can rearrange tis equation to read, $A = SDS^{-1}$. Also, the proof of Theorem DC [1039] says that the diagonal elements of D are the eigenvalues of A and we have the flexibility to assume they lie on the diagonal in the same order as we have specified above. Now, let $X^* = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$ be the columns of S, and let $Y^* = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_n\}$ be the rows of S^{-1} converted to column vectors. With little motivation other than the statement of the theorem, define size n matrices A_k , $1 \leq k \leq n$ by $A_k = \lambda_k \mathbf{x}_k \mathbf{y}_k^t$. Finally, let D_k be the size n matrix that is totally zero, other than having λ_k in row k and column k.

With everything in place, we compute entry-by-entry,

$$[A]_{ij} = [SDS^{-1}]_{ij}$$
 Definition DZM [1037]

 \square

$$= \left[S\left(\sum_{k=1}^{n} D_{k}\right) S^{-1} \right]_{ij}$$

$$= \left[S\left(\sum_{k=1}^{n} D_{k} S^{-1}\right) \right]_{ij}$$

$$= \left[\sum_{k=1}^{n} S D_{k} S^{-1} \right]_{ij}$$

$$= \sum_{k=1}^{n} \left[S D_{k} S^{-1} \right]_{ij}$$

$$= \sum_{k=1}^{n} \sum_{\ell=1}^{n} \left[S D_{k} \right]_{i\ell} \left[S^{-1} \right]_{\ell j}$$

$$= \sum_{k=1}^{n} \sum_{\ell=1}^{n} \sum_{p=1}^{n} \left[S \right]_{ip} \left[D_{k} \right]_{p\ell} \left[S^{-1} \right]_{\ell j}$$

$$= \sum_{k=1}^{n} \left[S \right]_{ik} \left[D_{k} \right]_{kk} \left[S^{-1} \right]_{kj}$$

$$= \sum_{k=1}^{n} \left[S \right]_{ik} \lambda_{k} \left[S^{-1} \right]_{kj}$$

Definition MA [427]

Theorem MMDAA [476]

Theorem MMDAA [476]

Definition MA [427]

Theorem EMP [471]

Theorem EMP [471]

$$[D_k]_{p\ell} = 0$$
 if $p \neq k$, or $\ell \neq k$

 $[D_k]_{kk} = \lambda_k$

$$= \sum_{k=1}^{n} \lambda_{k} [S]_{ik} [S^{-1}]_{kj}$$

$$= \sum_{k=1}^{n} \lambda_{k} [\mathbf{x}_{k}]_{i1} [\mathbf{y}_{k}^{t}]_{1j}$$

$$= \sum_{k=1}^{n} \lambda_{k} \sum_{q=1}^{1} [\mathbf{x}_{k}]_{iq} [\mathbf{y}_{k}^{t}]_{qj}$$

$$= \sum_{k=1}^{n} \lambda_{k} [\mathbf{x}_{k}\mathbf{y}_{k}^{t}]_{ij}$$

$$= \sum_{k=1}^{n} [\lambda_{k}\mathbf{x}_{k}\mathbf{y}_{k}^{t}]_{ij}$$

$$= \sum_{k=1}^{n} [A_{k}]_{ij}$$

Property CMCN [1597]

Definition of X^*, Y^*

Theorem EMP [471]

Definition MSM [428]

Definition of A_k

Definition MA [427]

So by Definition ME [427] we have the desired equality of matrices. The careful reader will have noted that $A_k = \mathcal{O}$, $r + 1 \le k \le n$, since $\lambda_k = 0$ in these instances. To get the sets

X and Y from X^* and Y^* , simply discard the last n - r vectors. We can safely ignore (or remove) $A_{r+1}, A_{r+2}, \ldots, A_n$ from the summation just derived.

One last assertion to check. What is the rank of A_k , $1 \le k \le r$? Every row of A_k is a scalar multiple of \mathbf{y}_k^t , row k of the nonsingular matrix S^{-1} (Theorem MIMI [521]). As a row of a nonsingular matrix, \mathbf{y}_k^t cannot be all zeros. In particular, row i of A_k is obtained as a scalar multiple of \mathbf{y}_k^t by the scalar $\alpha_k [\mathbf{x}_k]_i$. We have restricted ourselves to the nonzero eigenvalues of A, and as S is nonsingular, some entry of \mathbf{x}_k is nonzero. This all implies that some row of A_k will be nonzero. Now consider row-reducing A_k . Swap the nonzero row up into row 1. Use scalar multiples of this row to zero out every other row. This leaves a single nonzero row in the reduced row-echelon form, so A_k has rank one.

We record two observations that was not stated in our theorem above. First, the vectors in X, chosen as columns of S, are eigenvectors of A. Second, the product of two vectors from X and Y in the opposite order, by which we mean $\mathbf{y}_i^t \mathbf{x}_j$, is the entry in row *i* and column *j* of the matrix product $S^{-1}S = I_n$ (Theorem EMP [471]). In particular,

$$\mathbf{y}_i^t \mathbf{x}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

We give two computational examples. One small, one a bit bigger.

Example ROD2 Rank one decomposition, size 2 Consider the 2×2 matrix,

$$A = \begin{bmatrix} -16 & -6\\ 45 & 17 \end{bmatrix}$$

By the techniques of Chapter E [944] we find the eigenvalues and eigenspaces,

$$\lambda_1 = 2$$
 $\mathcal{E}_A(2) = \left\langle \left\{ \begin{bmatrix} -1\\ 3 \end{bmatrix} \right\} \right\rangle$ $\lambda_2 = -1$ $\mathcal{E}_A(-1) = \left\langle \left\{ \begin{bmatrix} -2\\ 5 \end{bmatrix} \right\} \right\rangle$

With n = 2 distinct eigenvalues, Theorem DED [1048] tells us that A is diagonalizable, and with no zero eigenvalues we see that A has full rank. Theorem DC [1039] says we can construct the nonsingular matrix S with eigenvectors of A as columns, so we have

$$S = \begin{bmatrix} -1 & -2\\ 3 & 5 \end{bmatrix} \qquad \qquad S^{-1} = \begin{bmatrix} 5 & 2\\ -3 & -1 \end{bmatrix}$$

From these matrices we obtain the sets of vectors

$$X = \left\{ \begin{bmatrix} -1\\3 \end{bmatrix}, \begin{bmatrix} -2\\5 \end{bmatrix} \right\} \qquad \qquad Y = \left\{ \begin{bmatrix} 5\\2 \end{bmatrix}, \begin{bmatrix} -3\\-1 \end{bmatrix} \right\}$$

And we have the matrices,

$$A_{1} = 2 \begin{bmatrix} -1\\ 3 \end{bmatrix} \begin{bmatrix} 5\\ 2 \end{bmatrix}^{t} = 2 \begin{bmatrix} -5 & -2\\ 15 & 6 \end{bmatrix} = \begin{bmatrix} -10 & -4\\ 30 & 12 \end{bmatrix}$$

$$A_{2} = (-1) \begin{bmatrix} -2\\5 \end{bmatrix} \begin{bmatrix} -3\\-1 \end{bmatrix}^{t} = (-1) \begin{bmatrix} 6 & 2\\-15 & -5 \end{bmatrix} = \begin{bmatrix} -6 & -2\\15 & 5 \end{bmatrix}$$

And you can easily verify that $A = A_1 + A_2$.

Here's a slightly larger example, and the matrix does not have full rank.

Example ROD4 Rank one decomposition, size 4

Consider the 4×4 matrix,

$$B = \begin{bmatrix} 34 & 18 & -1 & -6 \\ -44 & -24 & -1 & 9 \\ 36 & 18 & -3 & -6 \\ 36 & 18 & -6 & -3 \end{bmatrix}$$

By the techniques of Chapter E [944] we find the eigenvalues and eigenvectors,

$$\lambda_{1} = 3 \qquad \qquad \mathcal{E}_{B}(3) = \left\langle \left\{ \begin{bmatrix} 1\\-2\\1\\-1 \end{bmatrix}, \begin{bmatrix} 1\\-1\\1\\2 \end{bmatrix} \right\} \right\rangle$$
$$\lambda_{2} = -2 \qquad \qquad \mathcal{E}_{B}(-2) = \left\langle \left\{ \begin{bmatrix} -1\\2\\0\\0 \end{bmatrix} \right\} \right\rangle$$

 \boxtimes

$$\lambda_3 = 0 \qquad \qquad \mathcal{E}_A(0) = \left\langle \left\{ \begin{bmatrix} 2\\ -3\\ 2\\ 2 \end{bmatrix} \right\} \right\rangle$$

The algebraic and geometric multiplicities of each eigenvalue are equal, so Theorem DMFE [1044] tells us that A is diagonalizable. With a single zero eigenvalue we see that A has rank 4 - 1 = 3. Theorem DC [1039] says we can construct the nonsingular matrix S with eigenvectors of A as columns, so we have

$$S = \begin{bmatrix} 1 & 1 & -1 & 2 \\ -2 & -1 & 2 & -3 \\ 1 & 1 & 0 & 2 \\ -1 & 2 & 0 & 2 \end{bmatrix} \qquad \qquad S^{-1} = \begin{bmatrix} 4 & 2 & 0 & -1 \\ 8 & 4 & -1 & -1 \\ -1 & 0 & 1 & 0 \\ -6 & -3 & 1 & 1 \end{bmatrix}$$

Since r = 3, we need only collect three vectors from each of these matrices,

$$X = \left\{ \begin{bmatrix} 1\\ -2\\ 1\\ -1 \end{bmatrix}, \begin{bmatrix} 1\\ -1\\ 1\\ 2 \end{bmatrix}, \begin{bmatrix} -1\\ 2\\ 0\\ 0 \end{bmatrix} \right\} \qquad Y = \left\{ \begin{bmatrix} 4\\ 2\\ 0\\ -1 \end{bmatrix}, \begin{bmatrix} 8\\ 4\\ -1\\ -1 \end{bmatrix}, \begin{bmatrix} -1\\ 0\\ 1\\ 0 \end{bmatrix} \right\}$$

And we obtain the matrices,

Then we verify that

$$B = B_1 + B_2 + B_3$$

$$= \begin{bmatrix} 12 & 6 & 0 & -3 \\ -24 & -12 & 0 & 6 \\ 12 & 6 & 0 & -3 \\ -12 & -6 & 0 & 3 \end{bmatrix} + \begin{bmatrix} 24 & 12 & -3 & -3 \\ -24 & -12 & 3 & 3 \\ 24 & 12 & -3 & -3 \\ 48 & 24 & -6 & -6 \end{bmatrix} + \begin{bmatrix} -2 & 0 & 2 & 0 \\ 4 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 34 & 18 & -1 & -6\\ -44 & -24 & -1 & 9\\ 36 & 18 & -3 & -6\\ 36 & 18 & -6 & -3 \end{bmatrix}$$

 \boxtimes

Section TD Triangular Decomposition

This Section is a Draft, Subject to Changes

Our next decomposition will break a square matrix into a product of two matrices, one lower triangular and the other upper triangular. So we will write A = LU, and hence many refer to this as **LU decomposition**. We will see that this decomposition is very easy to compute and that it has a direct application to solving systems of equations. Since this section is about triangular matrices you might want to review the definitions and a couple of basic theorems back in Subsection OD.TM [1415].

Subsection TD Triangular Decomposition

With a slight condition on the nonsingularity of certain submatrices, we can split a matrix into a product of two triangular matrices.

Theorem TD

Triangular Decomposition

Suppose A is a square matrix of size n. Let A_k be the $k \times k$ matrix formed from A by taking the first k rows and the first k columns. Suppose that A_k is nonsingular for all $1 \le k \le n$. Then there is a lower triangular matrix L with all of its diagonal entries equal to 1 and an upper triangular matrix U such that A = LU. Furthermore, this decomposition is unique. \Box

Proof We will row reduce A to a row-equivalent upper triangular matrix through a series of row operations, forming intermediate matrices A'_{j} , $1 \leq j \leq n$, that denote the state of the conversion after working on column j. First, the lone entry of A_1 is $[A]_{11}$ and this scalar must be nonzero if A_1 is nonsingular (Theorem SMZD [935]). We can use row operations Definition RO [59] of the form $\alpha R_1 + R_k$, $2 \leq k \leq n$, where $\alpha = -[A]_{1k}/[A]_{11}$ to place zeros in the first column below the diagonal. The first two rows and columns of A'_1 are a 2×2 upper triangular matrix whose determinant is equal to the determinant of A_2 , since the matrices are row-equivalent through a sequence of row operations strictly of the third type (Theorem DRCMA [924]). As such the diagonal entries of this 2×2 submatrix of A'_1 are nonzero. We can employ this nonzero diagonal element with row operations of the form $\alpha R_2 + R_k$, $3 \leq k \leq n$ to place zeros below the diagonal in the second column. We can continue this process, column by column. The key observations are that our hypothesis on the nonsingularity of the A_k will guarantee a nonzero diagonal entry for each column when we need it, that the row operations employed are always of the third type using a multiple of a row to transform another row with a greater row index, and that the final result will be a nonsingular upper triangular matrix. This is the desired matrix U.

Each row operation described in the previous paragraph can be accomplished with matrix multiplication by the appropriate elementary matrix (Theorem EMDRO [886]). Since every row operation employed is adding a multiple of a row to a subsequent row these elementary matrices are of the form $E_{j,k}(\alpha)$ with j < k. By Definition ELEM [883], these matrices are lower triangular with every diagonal entry equal to 1. We know that the product of two such matrices will again be lower triangular (Theorem PTMT [1415]), but also, as you can also easily check using a proof with a style similar to one above, that the product maintains all 1's on the diagonal. Let $E_1, E_2, E_3, \ldots, E_m$ denote the elementary matrices for this sequence of row operations. Then

$$U = E_m E_{m-1} \dots E_3 E_2 E_1 A = L' A$$

where L' is the product of the elementary matrices, and we know L' is lower triangular with all 1's on the diagonal. Our desired matrix L is then $L = (L')^{-1}$. By Theorem ITMT [1417], L is lower triangular with all 1's on the diagonal and A = LU, as desired.

The process just described is deterministic. That is, the proof is constructive, with no freedom for each of us to walk through it differently. But could there be other matrices with the same properties as L and U that give such a decomposition of A. In other words, is the decomposition unique (Technique U [1625])? Suppose that we have two triangular decompositions, $A = L_1U_1$ and $A = L_2U_2$. Since A is nonsingular, two applications of Theorem NPNT [538] imply that L_1 , L_2 , U_1 , U_2 are all nonsingular. We have

$$L_2^{-1}L_1 = L_2^{-1}I_nL_1$$
Theorem MMIM [475]

$$= L_2^{-1}AA^{-1}L_1$$
Definition MI [506]

$$= L_2^{-1} L_2 U_2 (L_1 U_1)^{-1} L_1$$

= $L_2^{-1} L_2 U_2 U_1^{-1} L_1^{-1} L_1$ Theorem SS [520]
= $I_n U_2 U_1^{-1} I_n$ Definition MI [506]
= $U_2 U_1^{-1}$ Theorem MMIM [475]

Theorem ITMT [1417] tells us that L_2^{-1} is lower triangular and has 1's as the diagonal entries. By Theorem PTMT [1415], the product $L_2^{-1}L_1$ is again lower triangular, and it is simple to check (as before) that the diagonal entries of the product are again all 1's. By the entirely similar process we can conclude that the product $U_2U_1^{-1}$ is upper triangular. Because these two products are equal, their common value is a matrix that is both lower triangular *and* upper triangular, with all 1's on the diagonal. The only matrix meeting these three requirements is the identity matrix (Definition IM [175]). So, we have,

$$I_n = L_2^{-1}L_1 \Rightarrow L_2 = L_1$$
 $I_n = U_2U_1^{-1} \Rightarrow U_1 = U_2$

which establishes the uniqueness of the decomposition.

Studying the proofs of some previous theorems will perhaps give you an idea for an approach to computing a triangular decomposition. In the proof of Theorem CINM [516] we augmented a nonsingular matrix with an identity matrix of the same size, and row-reduced until the original matrix became the identity matrix (as we knew in advance would happen, since we knew Theorem NMRRI [176]). Theorem PEEF [621] tells us about properties of extended echelon form, and in particular, that B = JA, where A is the matrix that begins on the left, and B is the reduced row-echelon form of A. The matrix J is the result on the

right side of the augmented matrix, which is the result of applying the same row operations to the identity matrix. We should recognize now that J is just the product of the elementary matrices (Subsection DM.EM [883]) that perform these row operations. Theorem ITMT [1417] used the extended echelon form to discern properties of the inverse of a triangular matrix. Theorem TD [1898] proves the existence of a triangular decomposition by applying specific row operations, and tracking the relevant elementary row operations. It is not a great leap to combine these observations into a computational procedure.

To find the triangular decomposition of A, augment A with the identity matrix of the same size and call this new $2n \times n$ matrix, M. Perform row operations on M that convert the first n columns to an upper triangular matrix. Do this using only row operations that add a scalar multiple of one row to another row with higher index (i.e. lower down). In this way, the last n columns of M will be converted into a lower triangular matrix with 1's on the diagonal (since M has 1's in these locations initially). We could think of this process as doing about half of the work required to compute the inverse of A. Take the first n columns of the row-equivalent version of M and call this matrix U. Take the final n columns of the row-equivalent version of M and call this matrix L'. Then by a proof employing elementary matrices, or a proof similar in spirit to the one used to prove Theorem PEEF [621], we arrive at a result similar to the second assertion of Theorem PEEF [621]. Namely, U = L'A. Multiplication on the left, by the inverse of L', will give us a decomposition of A (which we know to be unique). Ready? Lets try it.

Example TD4

Triangular decomposition, size 4

In this example, we will illustrate the process for computing a triangular decomposition, as described in the previous paragraphs. Consider the nonsingular square matrix A of size 4,

$$A = \begin{bmatrix} -2 & 6 & -8 & 7\\ -4 & 16 & -14 & 15\\ -6 & 22 & -23 & 26\\ -6 & 26 & -18 & 17 \end{bmatrix}$$

We form M by augmenting A with the size 4 identity matrix I_4 . We will perform the allowed operations, column by column, only reporting intermediate results as we finish converting each column. It is easy to determine exactly which row operations we perform, since the final four columns contain a record of each such operation. We will not verify our hypotheses about the nonsingularity of the A_k , since if we do not have these conditions, we will reach a stage where a diagonal entry is zero and we cannot create the row operations we need to zero out the bottom portion of the associated column. In other words, we can boldly proceed and the necessity of our hypotheses will become apparent.

$$M = \begin{bmatrix} -2 & 6 & -8 & 7 & 1 & 0 & 0 & 0 \\ -4 & 16 & -14 & 15 & 0 & 1 & 0 & 0 \\ -6 & 22 & -23 & 26 & 0 & 0 & 1 & 0 \\ -6 & 26 & -18 & 17 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} -2 & 6 & -8 & 7 & 1 & 0 & 0 & 0 \\ 0 & 4 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 4 & 1 & 5 & -3 & 0 & 1 & 0 \\ 0 & 8 & 6 & -4 & -3 & 0 & 0 & 1 \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} -2 & 6 & -8 & 7 & 1 & 0 & 0 & 0 \\ 0 & 4 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & -1 & 4 & -1 & -1 & 1 & 0 \\ 0 & 0 & 2 & -6 & 1 & -2 & 0 & 1 \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} -2 & 6 & -8 & 7 & 1 & 0 & 0 & 0 \\ 0 & 4 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 4 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & -1 & 4 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 2 & -1 & -4 & 2 & 1 \end{bmatrix}$$

So at this point, we have U and L',

$$U = \begin{bmatrix} -2 & 6 & -8 & 7\\ 0 & 4 & 2 & 1\\ 0 & 0 & -1 & 4\\ 0 & 0 & 0 & 2 \end{bmatrix} \qquad \qquad L' = \begin{bmatrix} 1 & 0 & 0 & 0\\ -2 & 1 & 0 & 0\\ -1 & -1 & 1 & 0\\ -1 & -4 & 2 & 1 \end{bmatrix}$$

Then by whatever procedure we like (such as Theorem CINM [516]), we find

$$L = (L')^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 1 & 1 & 0 \\ 3 & 2 & -2 & 1 \end{bmatrix}$$

It is instructive to verify that indeed LU = A.

Subsection TDSSE Triangular Decomposition and Solving Systems of Equations

In this section we give an explanation of why you might be interested in a triangular decomposition for a matrix. Many of the computational problems in linear algebra revolve around solving large systems of equations, or nearly equivalently, finding inverses of large matrices. Suppose we have a system of equations with coefficient matrix A and vector of constants \mathbf{b} , and suppose further that A has the triangular decomposition A = LU.

Let \mathbf{y} be the solution to the linear system $\mathcal{LS}(L, \mathbf{b})$, so that by Theorem SLEMM [462], we have $L\mathbf{y} = \mathbf{b}$. Notice that since L is nonsingular, this solution is unique, and the form of L makes it trivial to solve the system. The first component of \mathbf{y} is determined easily, and we can continue on through determining the components of \mathbf{y} , without even ever dividing. Now,

 \boxtimes

with \mathbf{y} in hand, consider the linear system, $\mathcal{LS}(U, \mathbf{y})$. Let \mathbf{x} be the unique solution to this system, so by Theorem SLEMM [462] we have $U\mathbf{x} = \mathbf{y}$. Notice that a system of equations with U as a coefficient matrix is also straightforward to solve, though we will compute the bottom entries of \mathbf{x} first, and we will need to divide. The upshot of all this is that \mathbf{x} is a solution to $\mathcal{LS}(A, \mathbf{b})$, as we now show,

$$A\mathbf{x} = LU\mathbf{x} = L\left(U\mathbf{x}\right) = L\mathbf{y} = \mathbf{b}$$

An application of Theorem SLEMM [462] demonstrates that \mathbf{x} is a solution to $\mathcal{LS}(A, \mathbf{b})$.

Example TDSSE

Triangular decomposition solves a system of equations

Here we illustrate the previous discussion, recycling the decomposition found previously in Example TD4 [1902]. Consider the linear system $\mathcal{LS}(A, \mathbf{b})$ with

$$A = \begin{bmatrix} -2 & 6 & -8 & 7\\ -4 & 16 & -14 & 15\\ -6 & 22 & -23 & 26\\ -6 & 26 & -18 & 17 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} -10\\ -2\\ -1\\ -8 \end{bmatrix}$$

First we solve the system $\mathcal{LS}(L, \mathbf{b})$ (see Example TD4 [1902] for L),

$$y_1 = -10$$

$$2y_1 + y_2 = -2$$

$$3y_1 + y_2 + y_3 = -1$$

$$3y_1 + 2y_2 - 2y_3 + y_4 = -8$$

Then

$$y_1 = -10$$

$$y_2 = -2 - 2y_1 = -2 - 2(-10) = 18$$

$$y_3 = -1 - 3y_1 - y_2 = -1 - 3(-10) - 18 = 11$$

$$y_4 = -8 - 3y_1 - 2y_2 + 2y_3 = -8 - 3(-10) - 2(18) + 2(11) = 8$$

 \mathbf{SO}

$$\mathbf{y} = \begin{bmatrix} -10\\ 18\\ 11\\ 8 \end{bmatrix}$$

Then we solve the system $\mathcal{LS}(U, \mathbf{y})$ (see Example TD4 [1902] for U),

$$-2x_1 + 6x_2 - 8x_3 + 7x_4 = -10$$
$$4x_2 + 2x_3 + x_4 = 18$$
$$-x_3 + 4x_4 = 11$$
$$2x_4 = 8$$

Then

$$x_4 = 8/2 = 4$$

$$x_{3} = (11 - 4x_{4})/(-1) = (11 - 4(4))/(-1) = 5$$

$$x_{2} = (18 - 2x_{3} - x_{4})/4 = (18 - 2(5) - 4)/4 = 1$$

$$x_{1} = (-10 - 6x_{2} + 8x_{3} - 7x_{4})/(-2) = (-10 - 6(1) + 8(5) - 7(4))/(-2) = 2$$

And so

$$\mathbf{x} = \begin{bmatrix} 4\\5\\1\\2 \end{bmatrix}$$

is the solution to $\mathcal{LS}(U, \mathbf{y})$ and consequently is the unique solution to $\mathcal{LS}(A, \mathbf{b})$, as you can easily verify.

Subsection CTD Computing Triangular Decompositions

It would be a simple matter to adjust the algorithm for converting a matrix to reduced rowechelon form and obtain an algorithm to compute the triangular decomposition of the matrix, along the lines of Example TD4 [1902] and the discussion preceding this example. However, it is possible to obtain relatively simple formulas for the entries of the decomposition, and if computed in the proper order, an implementation will be straightforward. We will state the result as a theorem and then give an example of its use.

Theorem TDEE

Triangular Decomposition, Entry by Entry

Suppose that A is a square matrix of size n with a triangular decomposition A = LU, where L is lower triangular with diagonal entries all equal to 1, and U is upper triangular. Then

$$\begin{split} [U]_{ij} &= [A]_{ij} - \sum_{k=1}^{i-1} [L]_{ik} [U]_{kj} & 1 \le i \le j \le n \\ [L]_{ij} &= \frac{1}{[U]_{jj}} \left([A]_{ij} - \sum_{k=1}^{j-1} [L]_{ik} [U]_{kj} \right) & 1 \le j < i \le n \end{split}$$

Proof Consider a single scalar product of an entry of L with an entry of U of the form $[L]_{ik}[U]_{kj}$. By Definition LTM [1415], if k > i then $[L]_{ik} = 0$, while Definition UTM [1415], says that if k > j then $[U]_{kj} = 0$. So we can combine these two facts to assert that if $k > \min(i, j)$, $[L]_{ik}[U]_{kj} = 0$ since at least one term of the product will be zero. Employing this observation,

$$[A]_{ij} = \sum_{k=1}^{n} [L]_{ik} [U]_{kj}$$
 Theorem EMP [471]

 \square

$$\begin{split} &= \sum_{k=1}^{\min(i,j)} [L]_{ik} [U]_{kj} \\ \text{Now, assume that } 1 \leq i \leq j \leq n, \\ &[U]_{ij} = [A]_{ij} - [A]_{ij} + [U]_{ij} \\ &= [A]_{ij} - \sum_{k=1}^{\min(i,j)} [L]_{ik} [U]_{kj} + [U]_{ij} \\ &= [A]_{ij} - \sum_{k=1}^{i} [L]_{ik} [U]_{kj} + [U]_{ij} \\ &= [A]_{ij} - \sum_{k=1}^{i-1} [L]_{ik} [U]_{kj} - [L]_{ii} [U]_{ij} + [U]_{ij} \\ &= [A]_{ij} - \sum_{k=1}^{i-1} [L]_{ik} [U]_{kj} - [U]_{ij} + [U]_{ij} \\ &= [A]_{ij} - \sum_{k=1}^{i-1} [L]_{ik} [U]_{kj} - [U]_{ij} + [U]_{ij} \end{split}$$

And for $1 \leq j < i \leq n$,

$$[L]_{ij} = \frac{1}{[U]_{jj}} \left([L]_{ij} [U]_{jj} \right)$$

$$= \frac{1}{[U]_{jj}} \left([A]_{ij} - [A]_{ij} + [L]_{ij} [U]_{jj} \right)$$

$$= \frac{1}{[U]_{jj}} \left([A]_{ij} - \sum_{k=1}^{\min(i,j)} [L]_{ik} [U]_{kj} + [L]_{ij} [U]_{jj} \right)$$

$$= \frac{1}{[U]_{jj}} \left([A]_{ij} - \sum_{k=1}^{j} [L]_{ik} [U]_{kj} + [L]_{ij} [U]_{jj} \right)$$

$$= \frac{1}{[U]_{jj}} \left([A]_{ij} - \sum_{k=1}^{j-1} [L]_{ik} [U]_{kj} - [L]_{ij} [U]_{jj} + [L]_{ij} [U]_{jj} \right)$$

$$= \frac{1}{[U]_{jj}} \left([A]_{ij} - \sum_{k=1}^{j-1} [L]_{ik} [U]_{kj} \right)$$

At first glance, these formulas may look exceedingly complex. Upon closer examination, it looks even worse. We have expressions for entries of U that depend on other entries of U and also on entries of L. But then the formula for entries of L depend on entries from L and entries from U. Do these formula have circular dependencies? Or perhaps equivalently, how do we get started? The key is to be organized about the computations and employ these two (similar) formulas in a specific order. First compute the first row of L, followed by the first column of U. Then the second row of L, followed by the second column of U. And so on.

In this way, all of the values required for each new entry will have already been computed previously.

Of course, the formula for entries of L require division by diagonal entries of U. These entries might be zero, but in this case A is nonsingular and does not have a triangular decomposition. So we need not check the hypothesis carefully and can launch into the arithmetic dictated by the formulas, confident that we will be reminded when a decomposition is not possible. Note that these formula give us all of the values that we need for the decomposition, since we require that L has 1's on the diagonal. If we replace the 1's on the diagonal of L by zeros, and add the matrix U, we get an $n \times n$ matrix containing all the information we need to resurrect the triangular decomposition. This is mostly a notational convenience, but it is a frequent way of presenting the information. We'll employ it in the next example.

Example TDEE6 Triangular decomposition, entry by entry, size 6

We illustrate the application of the formulas in Theorem TDEE [1908] for the 6×6 matrix A.

$$A = \begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -6 & -4 & 5 & 2 & 4 & 2 \\ 9 & 9 & -7 & -7 & 0 & 1 \\ -6 & -10 & 8 & 10 & -1 & -7 \\ 6 & 4 & -9 & -2 & -10 & 1 \\ 9 & 3 & -12 & -3 & -21 & -2 \end{bmatrix}$$

Using the notational convenience of packaging the two triangular matrices into one matrix,

and using the ordering of the computations mentioned above, we display the results after computing a single row and column of each of the two triangular matrices.

$\begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -2 & & & & \\ 3 & & & & & \\ -2 & & & & \\ 2 & & & & & \\ 3 & & & & & \end{bmatrix}$	$\begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -2 & 2 & -1 & -2 & 2 & 2 \\ 3 & 0 & & & & \\ -2 & -2 & & & & \\ 2 & -1 & & & & \\ 3 & -3 & & & & \end{bmatrix}$
$\begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -2 & 2 & -1 & -2 & 2 & 2 \\ 3 & 0 & 2 & -1 & 3 & 1 \\ -2 & -2 & 0 & & & \\ 2 & -1 & -2 & & & \\ 3 & -3 & -3 & & & \end{bmatrix}$	$\begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -2 & 2 & -1 & -2 & 2 & 2 \\ 3 & 0 & 2 & -1 & 3 & 1 \\ -2 & -2 & 0 & 2 & 1 & -3 \\ 2 & -1 & -2 & -1 & & \\ 3 & -3 & -3 & -3 & & \end{bmatrix}$
$\begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -2 & 2 & -1 & -2 & 2 & 2 \\ 3 & 0 & 2 & -1 & 3 & 1 \\ -2 & -2 & 0 & 2 & 1 & -3 \\ 2 & -1 & -2 & -1 & 1 & 2 \\ 3 & -3 & -3 & -3 & 0 \end{bmatrix}$	$\begin{bmatrix} 3 & 3 & -3 & -2 & -1 & 0 \\ -2 & 2 & -1 & -2 & 2 & 2 \\ 3 & 0 & 2 & -1 & 3 & 1 \\ -2 & -2 & 0 & 2 & 1 & -3 \\ 2 & -1 & -2 & -1 & 1 & 2 \\ 3 & -3 & -3 & -3 & 0 & -2 \end{bmatrix}$

Splitting out the pieces of this matrix, we have the decomposition,

L =	$ \begin{array}{c} -2 \\ 3 \\ -2 \end{array} $	$\begin{array}{c} 0 \\ 1 \\ 0 \\ -2 \\ -1 \end{array}$	0 1 0	$egin{array}{c} 0 \ 0 \ 1 \end{array}$	0 0 0	0 0 0	U =	0 0 0 0	2 0 0 0	$-1 \\ 2 \\ 0 \\ 0 \\ 0$	0	2 3 1 1	$ \begin{array}{c} 2 \\ 1 \\ -3 \\ 2 \end{array} $	
	2		-2	-1	1	0		0	0	0		1	2	

 \boxtimes

The hypotheses of Theorem TD [1898] can be weakened slightly to include matrices where not every A_k is nonsingular. The introduces a rearrangement of the rows and columns of Ato force as many as possible of the smaller submatrices to be nonsingular. Then permutation matrices also enter into the decomposition. We will not present the details here, but instead suggest consulting a more advanced text on matrix analysis.

Section SVD Singular Value Decomposition

This Section is a Draft, Subject to Changes Needs Numerical Examples

The singular value decomposition is one of the more useful ways to represent any matrix, even rectangular ones. We can also view the singular values of a (rectangular) matrix as analogues of the eigenvalues of a square matrix. Our definitions and theorems in this section rely heavily on the properties of the matrix-adjoint products (A^*A and AA^*), which we first met in Theorem CPSM [1879]. We start by examining some of the basic properties of these two matrices. Now would be a good time to review the basic facts about positive semi-definite matrices in Section PSM [1878].

Subsection MAP Matrix-Adjoint Product

Theorem EEMAP

Eigenvalues and Eigenvectors of Matrix-Adjoint Product

Suppose that A is an $m \times n$ matrix and A^*A has rank r. Let $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p$ be the nonzero distinct eigenvalues of A^*A and let $\rho_1, \rho_2, \rho_3, \ldots, \rho_q$ be the nonzero distinct eigenvalues of AA^* . Then,

- 1. p = q.
- 2. The distinct nonzero eigenvalues can be ordered such that $\lambda_i = \rho_i, 1 \leq i \leq p$.
- 3. Properly ordered, $\alpha_{A^*A}(\lambda_i) = \alpha_{AA^*}(\rho_i), 1 \le i \le p$.
- 4. The rank of A^*A is equal to the rank of AA^* .
- 5. There is an orthonormal basis, $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$ of \mathbb{C}^n composed of eigenvectors of A^*A and an orthonormal basis, $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_m\}$ of \mathbb{C}^m composed of eigenvectors of AA^* with the following properties. Order the eigenvectors so that $\mathbf{x}_i, r+1 \leq i \leq n$ are the eigenvectors of A^*A for the zero eigenvalue. Let $\delta_i, 1 \leq i \leq r$ denote the nonzero eigenvalues of A^*A . Then $A\mathbf{x}_i = \sqrt{\delta_i}\mathbf{y}_i, 1 \leq i \leq r$ and $A\mathbf{x}_i = \mathbf{0}, r+1 \leq i \leq n$. Finally, $\mathbf{y}_i, r+1 \leq i \leq m$, are eigenvectors of AA^* for the zero eigenvalue.

 \square

Proof Suppose that $\mathbf{x} \in \mathbb{C}^n$ is any eigenvector of A^*A for a nonzero eigenvalue λ . We will show that $A\mathbf{x}$ is an eigenvector of AA^* for the same eigenvalue, λ . First, we ascertain that $A\mathbf{x}$ is not the zero vector.

$\langle A\mathbf{x}, A\mathbf{x} \rangle = \langle A\mathbf{x}, (A^*)^* \mathbf{x} \rangle$	Theorem AA $[443]$
$=\langle A^{*}A{f x},{f x} angle$	Theorem AIP [485]
$=\langle\lambda{f x},{f x} angle$	Definition EEM [945]
$=\lambda\left\langle \mathbf{x,x} ight angle$	Theorem IPSM [401]

Since **x** is an eigenvector, $\mathbf{x} \neq \mathbf{0}$, and by Theorem PIP [406], $\langle \mathbf{x}, \mathbf{x} \rangle \neq 0$. As λ was assumed to be nonzero, we see that $\langle A\mathbf{x}, A\mathbf{x} \rangle \neq 0$. Again, Theorem PIP [406] tells us that $A\mathbf{x} \neq \mathbf{0}$.

Much of the sequel turns on the following simple computation. If you ever wonder what all the fuss is about adjoints, Hermitian matrices, square roots, and singular values, return to this brief computation, as it holds the key. There is much more to do in this proof, but after this it is mostly bookkeeping. Here we go. We check that $A\mathbf{x}$ functions as an eigenvector of AA^* for the eigenvalue λ ,

$(AA^*) A\mathbf{x} = A (A^*A) \mathbf{x}$	Theorem MMA [478]
$=A\lambda\mathbf{x}$	Definition EEM [945]
$=\lambda\left(A\mathbf{x} ight)$	Theorem MMSMM [477]

That's it. If \mathbf{x} is an eigenvector of A^*A (for a *nonzero* eigenvalue), then $A\mathbf{x}$ is an eigenvector for AA^* for the same eigenvalue. Let's see what this buys us.

 A^*A and AA^* are Hermitian matrices (Definition HM [486]), and hence are normal (Definition NRML [1427]). This provides the existence of orthonormal bases of eigenvectors for each matrix by Theorem OBNM [1434]. Also, since each matrix is diagonalizable (Definition DZM [1037]) by Theorem OD [1429] we can interchange algebraic and geometric multiplicities by Theorem DMFE [1044].

Our first step is to establish that an eigenvalue λ has the same geometric multiplicity for both A^*A and AA^* . Suppose $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_s\}$ is an orthonormal basis of eigenvectors of A^*A for the eigenspace $\mathcal{E}_{A^*A}(\lambda)$. Then for $1 \leq i < j \leq s$, note

$\langle A\mathbf{x}_i, A\mathbf{x}_j \rangle = \langle A\mathbf{x}_i, (A^*)^* \mathbf{x}_j \rangle$	Theorem AA $[443]$
$=\langle A^*A{f x}_i,{f x}_j angle$	Theorem AIP [485]
$=\langle\lambda {f x}_i,{f x}_j angle$	Definition EEM [945]
$=\lambda\left\langle \mathbf{x}_{i},\mathbf{x}_{j} ight angle$	Theorem IPSM $[401]$
$=\lambda(0)$	Definition ONS [418]
= 0	Property ZCN [1597]

Then the set $E = \{A\mathbf{x}_1, A\mathbf{x}_2, A\mathbf{x}_3, \ldots, A\mathbf{x}_s\}$ is an orthogonal set of nonzero eigenvectors of AA^* for the eigenvalue λ . By Theorem OSLI [412], the set E is linearly independent and so the geometric multiplicity of λ as an eigenvalue of AA^* is s or greater. We have

$$\alpha_{A^*A}(\lambda) = \gamma_{A^*A}(\lambda) \le \gamma_{AA^*}(\lambda) = \alpha_{AA^*}(\lambda)$$

This inequality applies to any matrix, so long as the eigenvalue is nonzero. We now apply it

to the matrix A^* ,

$$\alpha_{AA^*}(\lambda) = \alpha_{(A^*)^*A^*}(\lambda) \le \alpha_{A^*(A^*)^*}(\lambda) = \alpha_{A^*A}(\lambda)$$

So for a nonzero eigenvalue, its algebraic multiplicities as an eigenvalue of A^*A and AA^* are equal. This is enough to establish that p = q and the eigenvalues can be ordered such that $\lambda_i = \rho_i$ for $1 \le i \le p$.

For any matrix B, the null space is identical to the eigenspace of the zero eigenvalue, $\mathcal{N}(B) = \mathcal{E}_B(0)$, and thus the nullity of the matrix is equal to the geometric multiplicity of the zero eigenvalue. With this, we can examine the ranks of A^*A and AA^* .

$$r(A^*A) = n - n(A^*A)$$

$$= \left(\alpha_{A^*A}(0) + \sum_{i=1}^p \alpha_{A^*A}(\lambda_i) \right) - n(A^*A)$$
Theorem NEM [1015]
$$= \left(\alpha_{A^*A}(0) + \sum_{i=1}^p \alpha_{A^*A}(\lambda_i) \right) - \gamma_{A^*A}(0)$$
Definition GME [969]
$$= \left(\alpha_{A^*A}(0) + \sum_{i=1}^p \alpha_{A^*A}(\lambda_i) \right) - \alpha_{A^*A}(0)$$
Theorem DMFE [1044]
$$= \sum_{i=1}^p \alpha_{A^*A}(\lambda_i)$$

$$= \sum_{i=1}^{p} \alpha_{AA^{*}} (\lambda_{i})$$

$$= \left(\alpha_{AA^{*}} (0) + \sum_{i=1}^{p} \alpha_{AA^{*}} (\lambda_{i}) \right) - \alpha_{AA^{*}} (0)$$

$$= \left(\alpha_{AA^{*}} (0) + \sum_{i=1}^{p} \alpha_{AA^{*}} (\lambda_{i}) \right) - \gamma_{AA^{*}} (0) \qquad \text{Theorem DMFE [1044]}$$

$$= \left(\alpha_{AA^{*}} (0) + \sum_{i=1}^{p} \alpha_{AA^{*}} (\lambda_{i}) \right) - n (AA^{*}) \qquad \text{Definition GME [969]}$$

$$= m - n (AA^{*}) \qquad \text{Theorem NEM [1015]}$$

$$= r (AA^{*}) \qquad \text{Theorem RPNC [831]}$$

When A is rectangular, the square matrices A^*A and AA^* have different sizes. With equal algebraic and geometric multiplicities for their common nonzero eigenvalues, the difference in their sizes is manifest in different algebraic multiplicities for the zero eigenvalue and different nullities. Specifically,

$$n(A^*A) = n - r \qquad \qquad n(AA^*) = m - r$$

Suppose that $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n$ is an orthonormal basis of \mathbb{C}^n composed of eigenvectors of A^*A and ordered so that $\mathbf{x}_i, r+1 \leq i \leq n$ are eigenvectors of AA^* for the zero eigenvalue. Denote the associated nonzero eigenvalues of A^*A for these eigenvectors by $\delta_i, 1 \leq i \leq r$. Then define

$$\mathbf{y}_i = \frac{1}{\sqrt{\delta_i}} A \mathbf{x}_i \qquad \qquad 1 \le i \le r$$

Let $\mathbf{y}_{r+1}, \mathbf{y}_{r+2}, \mathbf{y}_{r+2}, \ldots, \mathbf{y}_m$ be an orthonormal basis for the eigenspace $\mathcal{E}_{AA^*}(0)$, whose existence is guaranteed by Theorem GSP [413]. As scalar multiples of demonstrated eigenvectors of AA^* , $\mathbf{y}_i, 1 \leq i \leq r$ are also eigenvectors of AA^* , and $\mathbf{y}_i, r+1 \leq i \leq n$ have been chosen as eigenvectors of AA^* . These eigenvectors also have norm 1, as we now show. For $1 \leq i \leq r$,

$$\begin{aligned} \|\mathbf{y}_i\| &= \left\| \frac{1}{\sqrt{\delta_i}} A \mathbf{x}_i \right\| \\ &= \sqrt{\left\langle \frac{1}{\sqrt{\delta_i}} A \mathbf{x}_i, \frac{1}{\sqrt{\delta_i}} A \mathbf{x}_i \right\rangle} \\ &= \sqrt{\frac{1}{\sqrt{\delta_i}} \frac{1}{\sqrt{\delta_i}}} \left\langle A \mathbf{x}_i, A \mathbf{x}_i \right\rangle} \\ &= \sqrt{\frac{1}{\sqrt{\delta_i}} \frac{1}{\sqrt{\delta_i}}} \left\langle A \mathbf{x}_i, A \mathbf{x}_i \right\rangle} \\ &= \frac{1}{\sqrt{\delta_i}} \sqrt{\left\langle A \mathbf{x}_i, A \mathbf{x}_i \right\rangle} \\ &= \frac{1}{\sqrt{\delta_i}} \sqrt{\left\langle A \mathbf{x}_i, (A^*)^* \mathbf{x}_i \right\rangle} \end{aligned}$$

Theorem IPN [406]

Theorem IPSM [401]

Theorem HMRE [1020]

Theorem AA [443]

$$= \frac{1}{\sqrt{\delta_i}} \sqrt{\langle A^* A \mathbf{x}_i, \mathbf{x}_i \rangle}$$
$$= \frac{1}{\sqrt{\delta_i}} \sqrt{\langle \delta_i \mathbf{x}_i, \mathbf{x}_i \rangle}$$
$$= \frac{1}{\sqrt{\delta_i}} \sqrt{\delta_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle}$$
$$= \frac{1}{\sqrt{\delta_i}} \sqrt{\delta_i(1)}$$
$$= 1$$

Theorem AIP [485]

Definition EEM [945]

Theorem IPSM [401]

Definition ONS [418]

For $r+1 \leq i \leq n$, the \mathbf{y}_i have been chosen to have norm 1.

Finally we check orthogonality. Consider two eigenvectors \mathbf{y}_i and \mathbf{y}_j with $1 \leq i < j \leq m$. If these two vectors have different eigenvalues, then Theorem HMOE [1022] establishes that the two eigenvectors are orthogonal. If the two eigenvectors have a zero eigenvalue, then they are orthogonal by the choice of the orthonormal basis of $\mathcal{E}_{AA^*}(0)$. If the two eigenvectors have identical, nonzero, eigenvalues, then

$$\begin{aligned} \langle \mathbf{y}_i, \, \mathbf{y}_j \rangle &= \left\langle \frac{1}{\sqrt{\delta_i}} A \mathbf{x}_i, \, \frac{1}{\sqrt{\delta_j}} A \mathbf{x}_j \right\rangle \\ &= \frac{1}{\sqrt{\delta_i}} \overline{\frac{1}{\sqrt{\delta_j}}} \left\langle A \mathbf{x}_i, \, A \mathbf{x}_j \right\rangle \\ &= \frac{1}{\sqrt{\delta_i \delta_j}} \left\langle A \mathbf{x}_i, \, A \mathbf{x}_j \right\rangle \end{aligned}$$

Theorem IPSM [401]

Theorem HMRE [1020]

$$= \frac{1}{\sqrt{\delta_i \delta_j}} \langle A \mathbf{x}_i, (A^*)^* \mathbf{x}_j \rangle$$

$$= \frac{1}{\sqrt{\delta_i \delta_j}} \langle A^* A \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$= \frac{1}{\sqrt{\delta_i \delta_j}} \langle \delta_i \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$= \frac{\delta_i}{\sqrt{\delta_i \delta_j}} \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

$$= \frac{\delta_i}{\sqrt{\delta_i \delta_j}} (0)$$

$$= 0$$

Theorem AA [443]

Theorem AIP [485]

Definition EEM [945]

Theorem IPSM [401]

Definition ONS [418]

So $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_m\}$ is an orthonormal set of eigenvectors for AA^* . The critical relationship between these two orthonormal bases is present by design. For $1 \le i \le r$,

$$A\mathbf{x}_i = \sqrt{\delta_i} \frac{1}{\sqrt{\delta_i}} A\mathbf{x}_i = \sqrt{\delta_i} \mathbf{y}_i$$

For $r+1 \leq i \leq n$ we have

$$\langle A\mathbf{x}_i, A\mathbf{x}_i \rangle = \langle A\mathbf{x}_i, (A^*)^* \mathbf{x}_i \rangle$$
 Theorem AA [443]
= $\langle A^* A \mathbf{x}_i, \mathbf{x}_i \rangle$ Theorem AIP [485]
= $\langle \mathbf{0}, \mathbf{x}_i \rangle$ Definition EEM [945]

= 0

Definition IP [398]

So by Theorem PIP [406], $A\mathbf{x}_i = \mathbf{0}$.

Subsection SVD Singular Value Decomposition

The square roots of the eigenvalues of A^*A (or almost equivalently, AA^* !) are known as the singular values of A. Here is the definition.

Definition SV Singular Values

Suppose A is an $m \times n$ matrix. If the eigenvalues of A^*A are $\delta_1, \delta_2, \delta_3, \ldots, \delta_n$, then the singular values of A are $\sqrt{\delta_1}, \sqrt{\delta_2}, \sqrt{\delta_3}, \ldots, \sqrt{\delta_n}$.

Theorem EEMAP [1915] is a total setup for the singular value decomposition. This remarkable theorem says that *any* matrix can be broken into a product of three matrices. Two are square, and unitary. In light of Theorem UMPIP [549], we can view these matrices as transforming vectors or coordinates in a rotational fashion. The middle matrix of this decomposition is rectangular, but is as close to being diagonal as a rectangular matrix can be. Viewed as a transformation, this matrix effects, reflections, contractions or expansions

along axes — it stretches vectors. So any matrix, viewed as a transformation is the product of a rotation, a stretch and a rotation.

The singular value theorem can also be viewed as an application of our most general statement about matrix representations of linear transformations relative to different bases. Theorem MRCB [1371] concerns linear transformations $T: U \to V$ where U and V are possibly different vector spaces. When U and V have different dimensions, the resulting matrix representation will be rectangular. In Section CB [1355] we quickly specialized to the case where U = V and the matrix representations are square with one of our most central results, Theorem SCB [1377]. Theorem SVD [1924] is an application of the full generality of Theorem MRCB [1371] where the relevant bases are now orthonormal sets.

Theorem SVD Singular Value Decomposition

Suppose A is an $m \times n$ matrix of rank r with nonzero singular values $s_1, s_2, s_3, \ldots, s_r$. Then $A = UDV^*$ where U is a unitary matrix of size m, V is a unitary matrix of size n and D is an $m \times n$ matrix given by

$$[D]_{ij} = \begin{cases} s_i & \text{if } 1 \le i = j \le r \\ 0 & \text{otherwise} \end{cases}$$

Proof Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_m$ be the orthonormal bases described by the conclusion of Theorem EEMAP [1915]. Define U to be the $m \times m$ matrix whose

 \square

columns are \mathbf{y}_i , $1 \leq i \leq m$, and define V to be the $n \times n$ matrix whose columns are \mathbf{x}_i , $1 \leq i \leq n$. With orthonormal sets of columns, by Theorem CUMOS [547] both U and V are unitary matrices.

Then for $1 \leq i \leq m, 1 \leq j \leq n$,

$$AV]_{ij} = [A\mathbf{x}_j]_i$$

= $\left[\sqrt{\delta_j}\mathbf{y}_j\right]_i$
= $[s_j\mathbf{y}_j]_i$
= $[\mathbf{y}_j]_i s_j$
= $[U]_{ij} [D]_{jj}$
= $\sum_{k=1}^m [U]_{ik} [D]_{kj}$
= $[UD]_{ij}$

Definition MM [468] Theorem EEMAP [1915] Definition SV [1923] Definition CVSM [203]

Theorem EMP [471]

So by Theorem ME [1016], AV = UD and thus

$$A = AI_n = AVV^* = UDV^*$$

Section SR Square Roots

This Section is a Draft, Subject to Changes Needs Numerical Examples

With all our results about Hermitian matrices, their eigenvalues and their diagonalizations, it will be a nearly trivial matter to now construct a "square root" of a positive semi-definite matrix. We will describe the square root of a matrix A as a matrix S such that $A = S^2$. In general, a matrix A might have many such square roots. But with a few results in hand we will be able to impose an extra condition on S that will make a unique S such that $A = S^2$. At that point we can define *the* square root of A formally.

Subsection SRM Square Root of a Matrix

Theorem **PSMSR**

Positive Semi-Definite Matrices and Square Roots

Suppose A is a square matrix. There is a positive semi-definite matrix S such that $A = S^2$ if and only if A is positive semi-definite.

Proof Let n denote the size of A.

(\Leftarrow) Suppose that A is positive semi-definite. Since A is Hermitian (Definition PSM [1879]) we know A is normal (Definition NRML [1427]) and so by Theorem OD [1429] there is a unitary matrix U and a diagonal matrix D, whose diagonal entries are the eigenvalues of A, such that $D = U^*AU$. The eigenvalues of A are all non-negative (Theorem EPSM [1880]), which allows us to define a diagonal matrix E whose diagonal entries are the positive square roots of the eigenvalues of A, in the same order as they appear in D. More precisely, define E to be the diagonal matrix with non-negative diagonal entries such that $E^2 = D$. Set $S = UEU^*$, and compute

$$S^{2} = UEU^{*}UEU^{*}$$

= $UEI_{n}EU^{*}$ Definition UM [544]
= $UEEU^{*}$ Theorem MMIM [475]

$$= UDU^{*}$$

$$= UU^{*}AUU^{*}$$
Theorem OD [1429]
$$= I_{n}AI_{n}$$
Definition UM [544]
$$= A$$
Theorem MMIM [475]

We need to first verify that S is Hermitian.

TTDTT

$$S^* = (UEU^*)^*$$

$$= (UEU^*)^*$$

$$= (U^*)^* E^*U^*$$
Theorem MMAD [483]

$$= UE^*U^*$$
Theorem AA [443]

$$= U(\overline{E})^t U^*$$
Definition A [442]

$$= UE^tU^*$$
Theorem HMRE [1020

$$= UEU^*$$
Diagonal matrix

$$= S$$

And finally, we want to check the use of S in an inner product. Notice that E is Hermitian since it is a diagonal matrix with real entries. Furthermore, as a diagonal matrix, the eigenvalues of E are precisely the diagonal entries, and since these were *chosen* to be positive, an application of Theorem EPSM [1880] tells us that E is positive semi-definite. Now, for any $\mathbf{x} \in \mathbb{C}^n$,

$$\langle S\mathbf{x}, \, \mathbf{x} \rangle = \langle UEU^*\mathbf{x}, \, \mathbf{x} \rangle$$

$$= \langle EU^* \mathbf{x}, U^* \mathbf{x} \rangle$$
Theorem AIP [485]
$$= \langle E(U^* \mathbf{x}), U^* \mathbf{x} \rangle$$

$$\geq 0$$
Definition PSM [1879]

So, according to Definition PSM [1879], S is positive semi-definite.

 (\Rightarrow) Assume that $A = S^2$, with S positive semi-definite. Then S is Hermitian, and we check that A is Hermitian.

$$A^* = (SS)^*$$

= S^*S^* Theorem MMAD [483]
= SS Definition HM [486]
= A

Now for the use of A in an inner product. For any $\mathbf{x} \in \mathbb{C}^n$,

So by Definition PSM [1879], A is positive semi-definite.

There is a very close relationship between the eigenvalues and eigenspaces of a positive semi-definite matrix and its positive semi-definite square root. The next theorem is interesting in its own right, but is also an important technical step in some other important results, such as the upcoming uniqueness of the square root (Theorem USR [1933]).

Theorem EESR

Eigenvalues and Eigenspaces of a Square Root

Suppose that A is a positive semi-definite matrix and S is a positive semi-definite matrix such that $A = S^2$. If $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p$ are the distinct eigenvalues of A, then the distinct eigenvalues of S are $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}, \ldots, \sqrt{\lambda_p}$, and $\mathcal{E}_S(\sqrt{\lambda_i}) = \mathcal{E}_A(\lambda_i)$ for $1 \le i \le p$. \Box

Proof Let **x** be an eigenvector of S for an eigenvalue ρ . Then, in the style of Theorem EPM [1007],

$$A\mathbf{x} = S^2 \mathbf{x} = S(S\mathbf{x}) = S(\rho \mathbf{x}) = \rho S \mathbf{x} = \rho^2 \mathbf{x}$$

so ρ^2 is an eigenvalue of A and must equal some λ_i . Furthermore, because S is positive semi-definite, Theorem EPSM [1880] tells us that $\rho \geq 0$. The impact for us here is that we cannot have two different eigenvalues of S whose squares equal the same eigenvalue of A, so we can pair each eigenvalue of S with a different eigenvalue of A, equal to its square. (A good exercise is to track through the rest of this proof in the situation where S is not assumed to be positive semi-definite and we do not have this condition on the eigenvalues. Where does the proof then break down?) Let ρ_i , $1 \leq i \leq q$ denote the q distinct eigenvalues of S. The discussion above implies that we can order the eigenvalues of A and S so that $\lambda_i = \rho_i^2$ for $1 \leq i \leq q$. Notice that at this point we know that $q \leq p$, though we will be showing that q = p. Additionally, the equation above tells us that every eigenvector of S for ρ_i is again an eigenvector of A for ρ_i^2 . So for $1 \le i \le q$, the relevant eigenspaces are related by

$$\mathcal{E}_{S}\left(\sqrt{\lambda_{i}}\right) = \mathcal{E}_{S}\left(\rho_{i}\right) \subseteq \mathcal{E}_{A}\left(\rho_{i}^{2}\right) = \mathcal{E}_{A}\left(\lambda_{i}\right)$$

So the eigenspaces of S are subsets of the eigenspaces of A, for the related eigenvalues. However, we will be showing that these sets are indeed equal to each other.

Both A and S are positive semi-definite, hence Hermitian and therefore normal. Theorem OD [1429] then tells us that each is diagonalizable (Definition DZM [1037]). Then Theorem DMFE [1044] says that the algebraic multiplicity and geometric multiplicity of each eigenvalue are equal. Then, if we let n denote the size of A,

$$n = \sum_{i=1}^{q} \alpha_{S} \left(\sqrt{\lambda_{i}} \right)$$
$$= \sum_{i=1}^{q} \gamma_{S} \left(\sqrt{\lambda_{i}} \right)$$
$$= \sum_{i=1}^{q} \dim \left(\mathcal{E}_{S} \left(\sqrt{\lambda_{i}} \right) \right)$$
$$\leq \sum_{i=1}^{q} \dim \left(\mathcal{E}_{A} \left(\lambda_{i} \right) \right)$$

Theorem NEM [1015]

Theorem DMFE [1044]

Definition GME [969]

Theorem PSSD [857]

$\leq \sum_{i=1}^{p} \dim \left(\mathcal{E}_{A} \left(\lambda_{i} ight) ight)$	Definition D [815]
$=\sum_{i=1}^{p}\gamma_{A}\left(\lambda_{i}\right)$	Definition GME [969]
$=\sum_{i=1}^{p}\alpha_{A}\left(\lambda_{i}\right)$	Theorem DMFE [1044]
= n	Theorem NEM [1015]

With equal values at the two ends of this chain of equalities and inequalities, we know that the two inequalities are forced to actually be equalities. In particular, the second inequality implies that p = q and the first, in conjunction with Theorem EDYES [858], implies that $\mathcal{E}_S(\sqrt{\lambda_i}) = \mathcal{E}_A(\lambda_i)$ for $1 \leq i \leq p$.

Notice that we defined the singular values of a matrix A as the square roots of the eigenvalues of A^*A (Definition SV [1923]). With Theorem EESR [1930] in hand we recognize the singular values of A as simply the eigenvalues of $A^*A^{1/2}$. Indeed, many authors take this as the definition of singular values, since it is equivalent to our definition. We have chosen not to wait for a discussion of square roots before making a definition of singular values, allowing us to present the singular value decomposition (Theorem SVD [1924]) all the sooner.

In the first half of the proof of Theorem PSMSR [1927] we could have chosen the matrix E (which was the essential component of the desired matrix S) in a variety of ways. Any collection of diagonal entries of E could be replaced by their negatives and we would maintain

the property that $E^2 = D$. However, if we decide to enforce the entries of E as non-negative quantities then E is positive semi-definite, and then S follows along as a positive semi-definite matrix. We now show that of all the possible square roots of a positive semi-definite matrix, only one is itself again positive semi-definite. In other words, the S of Theorem PSMSR [1927] is unique.

Theorem USR

Unique Square Root

Suppose A is a positive semi-definite matrix. Then there is a unique positive semi-definite matrix S such that $A = S^2$.

Proof Theorem PSMSR [1927] gives us the existence of at least one positive semi-definite matrix S such that $A = S^2$. As usual, we will assume that S_1 and S_2 are positive semi-definite matrices such that $A = S_1^2 = S_2^2$ (Technique U [1625]).

As A is diagonalizable, there is a basis of \mathbb{C}^n composed entirely of eigenvectors of A (Theorem DC [1039]), say $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$. Let $\delta_1, \delta_2, \delta_3, \dots, \delta_n$ denote the associated eigenvalues. Theorem EESR [1930] allows to conclude that $\mathcal{E}_A(\delta_i) = \mathcal{E}_{S_1}(\sqrt{\delta_i}) = \mathcal{E}_{S_2}(\sqrt{\delta_i})$. So $S_1\mathbf{x}_i = \sqrt{\delta_i}\mathbf{x}_i = S_2\mathbf{x}_i$ for $1 \le i \le n$.

Choose any $\mathbf{x} \in \mathbb{C}^n$. The spanning property of *B* allows us to conclude the existence of a set of scalars, $a_1, a_2, a_3, \ldots, a_n$, yielding \mathbf{x} as a linear combination of the vectors in *B*. So,

$$S_1 \mathbf{x} = S_1 \sum_{i=1}^n a_i \mathbf{x}_i = \sum_{i=1}^n a_i S_1 \mathbf{x}_i = \sum_{i=1}^n a_i \sqrt{\delta_i} \mathbf{x}_i = \sum_{i=1}^n a_i S_2 \mathbf{x}_i = S_2 \sum_{i=1}^n a_i \mathbf{x}_i = S_2 \mathbf{x}_i$$

Since S_1 and S_2 have the same action on every vector, Theorem EMMVP [466] yields the conclusion that $S_1 = S_2$.

With a criteria that distinguishes one square root from all the rest (positive semidefiniteness) we can now define *the* square root of a positive semi-definite matrix.

Definition SRM

Square Root of a Matrix

Suppose A is a positive semi-definite matrix and S is the positive semi-definite matrix such that $S^2 = SS = A$. Then S is the **square root** of A and we write $S = A^{1/2}$. (This definition contains Notation SRM.)

Section POD Polar Decomposition

This Section is a Draft, Subject to Changes Needs Numerical Examples

The polar decomposition of a matrix writes any matrix as the product of a unitary matrix (Definition UM [544]) and a positive semi-definite matrix (Definition PSM [1879]). It takes its name from a special way to write complex numbers. If you've had a basic course in complex analysis, the next paragraph will help explain the name. If the next paragraph makes no sense to you, there's no harm in skipping it.

Any complex number $z \in \mathbb{C}$ can be written as $z = re^{i\theta}$ where r is a positive number (computed as a square root of a function of the real and imaginary parts of z) and θ is an angle of rotation that converts 1 to the complex number $e^{i\theta} = \cos(\theta) + i\sin(\theta)$. The polar form of a square matrix is a product of a positive semi-definite matrix that is a square root of a function of the matrix together with a unitary matrix, which can be viewed as achieving a rotation (Theorem UMPIP [549]).

OK, enough preliminaries. We have all the tools in place to jump straight to our main theorem.

Theorem PDM

Polar Decomposition of a Matrix

Suppose that A is a square matrix. Then there is a unitary matrix U such that $A = (AA^*)^{1/2} U$.

Proof This theorem only claims the existence of a unitary matrix U that does a certain job. We will manufacture U and check that it meets the requirements.

Suppose A has size n and rank r. We begin by applying Theorem EEMAP [1915] to A. Let $B = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$ be the orthonormal basis of \mathbb{C}^n composed of eigenvectors for A^*A , and let $C = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_n\}$ be the orthonormal basis of \mathbb{C}^n composed of eigenvectors for AA^* . We have $A\mathbf{x}_i = \sqrt{\delta_i}\mathbf{x}_i$, $1 \le i \le r$, and $A\mathbf{x}_i = \mathbf{0}$, $r+1 \le i \le n$, where δ_i , $1 \le i \le r$ are the distinct nonzero eigenvalues of A^*A .

Define $T: \mathbb{C}^n \to \mathbb{C}^n$ to be the unique linear transformation such that $T(\mathbf{x}_i) = \mathbf{y}_i$, $1 \leq i \leq n$, as guaranteed by Theorem LTDB [1094]. Let E be the basis of standard unit vectors for \mathbb{C}^n (Definition SUV [409]), and define U to be the matrix representation (Definition MR [1286]) of T with respect to E, more carefully $U = M_{E,E}^T$. This is the matrix we are after. Notice that

$U\mathbf{x}_{i} = M_{E,E}^{T}\rho_{E}\left(\mathbf{x}_{i}\right)$	Definition VR $[1261]$
$=\rho_{E}\left(T\left(\mathbf{x}_{i}\right)\right)$	Theorem FTMR [1293]
$=\rho_{E}\left(\mathbf{y}_{i}\right)$	Theorem FTMR [1293]
$=\mathbf{y}_i$	Definition VR [1261]

Since B and C are orthonormal bases, and C is the result of multiplying the vectors of B by U, we conclude that U is unitary by Theorem UMCOB [798]. So once again, Theorem

EEMAP [1915] is a big part of the setup for a decomposition.

Let $\mathbf{x} \in \mathbb{C}^n$ be any vector. Since B is a basis of \mathbb{C}^n , there are scalars $a_1, a_2, a_3, \ldots, a_n$ expressing \mathbf{x} as a linear combination of the vectors in B. then

$$AA^{*})^{1/2} U\mathbf{x} = (AA^{*})^{1/2} U \sum_{i=1}^{n} a_{i} \mathbf{x}_{i}$$
Definition B [777]

$$= \sum_{i=1}^{n} (AA^{*})^{1/2} Ua_{i} \mathbf{x}_{i}$$
Theorem MMDAA [476]

$$= \sum_{i=1}^{n} a_{i} (AA^{*})^{1/2} U\mathbf{x}_{i}$$
Theorem MMSMM [477]

$$= \sum_{i=1}^{n} a_{i} (AA^{*})^{1/2} \mathbf{y}_{i}$$

$$= \sum_{i=1}^{r} a_{i} (AA^{*})^{1/2} \mathbf{y}_{i} + \sum_{i=r+1}^{n} a_{i} (AA^{*})^{1/2} \mathbf{y}_{i}$$
Property AAC [205]

$$= \sum_{i=1}^{r} a_{i} \sqrt{\delta_{i}} \mathbf{y}_{i} + \sum_{i=r+1}^{n} a_{i} (0) \mathbf{y}_{i}$$
Theorem EESR [1930]

$$= \sum_{i=1}^{r} a_{i} \sqrt{\delta_{i}} \mathbf{y}_{i} + \sum_{i=r+1}^{n} a_{i} \mathbf{0}$$
Theorem ZSSM [676]

$$= \sum_{i=1}^{r} a_i A \mathbf{x}_i + \sum_{i=r+1}^{n} a_i A \mathbf{x}_i$$
 Theorem EEMAP [1915]
$$= \sum_{i=1}^{n} a_i A \mathbf{x}_i$$
 Property AAC [205]
$$= \sum_{i=1}^{n} A a_i \mathbf{x}_i$$
 Theorem MMSMM [477]
$$= A \sum_{i=1}^{n} a_i \mathbf{x}_i$$
 Theorem MMDAA [476]
$$= A \mathbf{x}$$

So by Theorem EMMVP [466] we have the matrix equality $(AA^*)^{1/2} U = A$.

Part A Applications

Section CF Curve Fitting

This Section is Incomplete

Given two points in the plane, there is a unique line through them. Given three points in the plane, and not in a line, there is a unique parabola through them. Given four points in the plane, there is a unique polynomial, of degree 3 or less, passing through them. And so on. We can prove this result, and give a procedure for finding the polynomial with the help of Vandermonde matrices (Section VM [1870]).

Theorem IP

Interpolating Polynomial

Suppose $\{(x_i, y_i) | 1 \le i \le n+1\}$ is a set of n+1 points in the plane where the x-coordinates are all different. Then there is a unique polynomial of degree n or less, p(x), such that $p(x_i) = y_i, 1 \le i \le n+1$.

Proof Write $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$. To meet the conclusion of the theorem, we desire,

$$y_i = p(x_i) = a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_n x_i^n \qquad 1 \le i \le n+1$$

This is a system of n + 1 linear equations in the n + 1 variables $a_0, a_1, a_2, \ldots, a_n$. The vector of constants in this system is the vector containing the *y*-coordinates of the points. More importantly, the coefficient matrix is a Vandermonde matrix (Definition VM [1870]) built from the *x*-coordinates $x_1, x_2, x_3, \ldots, x_{n+1}$. Since we have required that these scalars all be different, Theorem NVM [1877] tells us that the coefficient matrix is nonsingular and Theorem NMUS [180] says the solution for the coefficients of the polynomial exists, and is unique. As a practical matter, Theorem SNCM [543] provides an expression for the solution.

Example PTFP Polynomial through five points

Suppose we have the following 5 points in the plane and we wish to pass a degree 4 polynomial through them.

i	1	2	3	4	5
x_i	-3	-1	2	3	6
y_i	276	16	31	$144 \ 2319$	

The required system of equations has a coefficient matrix that is the Vandermonde matrix where row i is successive powers of x_i

$$A = \begin{bmatrix} 1 & -3 & 9 & -27 & 81 \\ 1 & -1 & 1 & -1 & 1 \\ 1 & 2 & 4 & 8 & 16 \\ 1 & 3 & 9 & 27 & 81 \\ 1 & 6 & 36 & 216 & 1296 \end{bmatrix}$$

Theorem NMUS [180] provides a solution as

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = A^{-1} \begin{bmatrix} 276 \\ 16 \\ 31 \\ 144 \\ 2319 \end{bmatrix} = \begin{bmatrix} -\frac{1}{15} & \frac{9}{14} & \frac{9}{10} & -\frac{1}{2} & \frac{1}{42} \\ 0 & -\frac{3}{7} & \frac{3}{4} & -\frac{1}{3} & \frac{1}{84} \\ \frac{5}{108} & -\frac{1}{56} & -\frac{1}{4} & \frac{17}{72} & -\frac{11}{756} \\ -\frac{1}{54} & \frac{1}{21} & -\frac{1}{12} & \frac{1}{18} & -\frac{1}{756} \\ \frac{1}{56} & -\frac{1}{2} & \frac{1}{216} & -\frac{1}{2} & \frac{1}{216} \end{bmatrix} \begin{bmatrix} 276 \\ 16 \\ 31 \\ 144 \\ 2319 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \\ 5 \\ -2 \\ 2 \end{bmatrix}$$

So the polynomial is $p(x) = 3 - 4x + 5x^2 - 2x^3 + 2x^4$.

The unique polynomial passing through a set of points is known as the **interpolating polynomial** and it has many uses. Unfortunately, when confronted with data from an experiment the situation may not be so simple or clear cut. Read on.

Subsection DF Data Fitting

Suppose that we have n real variables, $x_1, x_2, x_3, \ldots, x_n$, that we can measure in an experiment. We believe that these variables combine, in a linear fashion, to equal another real variable, y. In other words, we have reason to believe from our understanding of the experiment, that

$$y = a_1 x_1 + a_2 x_2 + a_3 x_3 + \dots + a_n x_n$$

 \boxtimes

where the scalars $a_1, a_2, a_3, \ldots, a_n$ are not known to us, but are instead desirable. We would call this our model of the situation. Then we run the experiment m times, collecting sets of values for the variables of the experiment. For run number k we might denote these values as $y_k, x_{k1}, x_{k2}, x_{k3}, \ldots, x_{kn}$. If we substitute these values into the model equation, we get m linear equations in the unknown coefficients $a_1, a_2, a_3, \ldots, a_n$. If m = n, then we have a square coefficient matrix of the system which might happen to be nonsingular and there would be a unique solution.

However, more likely m > n (the more data we collect, the greater our confidence in the results) and the resulting system is inconsistent. It may be that our model is only an approximate understanding of the relationship between the x_i and y, or our measurements are not completely accurate. Still we would like to understand the situation we are studying, and would like some best answer for $a_1, a_2, a_3, \ldots, a_n$.

Let **y** denote the vector with $[\mathbf{y}]_i = y_i$, $1 \le i \le m$, let **a** denote the vector with $[\mathbf{a}]_j = a_j$, $1 \le j \le n$, and let X denote the $m \times n$ matrix with $[X]_{ij} = x_{ij}$, $1 \le i \le m$, $1 \le j \le n$. Then the model equation, evaluated with each run of the experiment, translates to $X\mathbf{a} = \mathbf{y}$. With the presumption that this system has no solution, we can try to minimize the difference between the two side of the equation $\mathbf{y} - X\mathbf{a}$. As a vector, it is hard to imagine what the minimum might be, so we instead minimize the square of its norm

$$S = (\mathbf{y} - X\mathbf{a})^t (\mathbf{y} - X\mathbf{a})$$

To keep the logical flow accurate, we will define the minimizing value and then give the proof that it behaves as desired.

Definition LSS

Least Squares Solution

Given the equation $X\mathbf{a} = \mathbf{y}$, where X is an $m \times n$ matrix of rank n, the least squares solution for \mathbf{a} is $(X^tX)^{-1}X^t\mathbf{y}$.

Theorem LSMR

Least Squares Minimizes Residuals

Suppose that X is an $m \times n$ matrix of rank n. The least squares solution of $X\mathbf{a} = \mathbf{y}$, $\mathbf{a}' = (X^t X)^{-1} X^t \mathbf{y}$, minimizes the expression

$$S = \left(\mathbf{y} - X\mathbf{a}\right)^t \left(\mathbf{y} - X\mathbf{a}\right)$$

Proof We begin by finding the critical points of S. In preparation, let \mathbf{X}_j denote column j of X, for $1 \leq j \leq n$ and compute partial derivatives with respect to a_j , $1 \leq j \leq n$. A matrix product of the form $\mathbf{x}^t \mathbf{y}$ is a sum of products, so a derivative is a sum of applications of the product rule,

$$\frac{\partial}{\partial a_j} S = \frac{\partial}{\partial a_j} \left((\mathbf{y} - X\mathbf{a})^t (\mathbf{y} - X\mathbf{a}) \right)$$
$$= \sum_{i=1}^m \frac{\partial}{\partial a_j} \left([\mathbf{y} - X\mathbf{a}]_i \right) [\mathbf{y} - X\mathbf{a}]_i + [\mathbf{y} - X\mathbf{a}]_i \frac{\partial}{\partial a_j} \left([\mathbf{y} - X\mathbf{a}]_i \right)$$
$$= 2\sum_{i=1}^m \frac{\partial}{\partial a_j} \left([\mathbf{y} - X\mathbf{a}]_i \right) [\mathbf{y} - X\mathbf{a}]_i$$

$$= 2 \sum_{i=1}^{m} \frac{\partial}{\partial a_{j}} \left([\mathbf{y}]_{i} - \sum_{k=1}^{n} [X]_{ik} [\mathbf{a}]_{k} \right) [\mathbf{y} - X\mathbf{a}]_{i}$$
$$= 2 \sum_{i=1}^{m} - [X]_{ij} [\mathbf{y} - X\mathbf{a}]_{i}$$
$$= -2 (\mathbf{X}_{j})^{t} (\mathbf{y} - X\mathbf{a})$$

The first partial derivatives will allow us to find critical points, while second partial derivatives will be needed to confirm that a critical point will yield a minimum. Return to the next-to-last expression for the first partial derivative of S,

$$\begin{aligned} \frac{\partial}{\partial a_{\ell} a_{j}} S &= \frac{\partial}{\partial a_{\ell}} 2 \sum_{i=1}^{m} - [X]_{ij} \left[\mathbf{y} - X \mathbf{a} \right]_{i} \\ &= -2 \sum_{i=1}^{m} \frac{\partial}{\partial a_{\ell}} \left[X \right]_{ij} \left[\mathbf{y} - X \mathbf{a} \right]_{i} \\ &= -2 \sum_{i=1}^{m} [X]_{ij} \frac{\partial}{\partial a_{\ell}} \left(\left[\mathbf{y} \right]_{i} - \sum_{k=1}^{n} [X]_{ik} \left[\mathbf{a} \right]_{k} \right) \\ &= -2 \sum_{i=1}^{m} [X]_{ij} \left(- [X]_{i\ell} \right) \end{aligned}$$

$$= 2 \sum_{i=1}^{m} [X]_{ij} [X]_{i\ell}$$
$$= 2 \sum_{i=1}^{m} [X^t]_{ji} [X]_{i\ell}$$
$$= 2 [X^t X]_{j\ell}$$

For $1 \leq j \leq n$, set $\frac{\partial}{\partial a_i}S = 0$. This results in the *n* scalar equations

$$(\mathbf{X}_j)^t X \mathbf{a} = (\mathbf{X}_j)^t \mathbf{y} \qquad 1 \le j \le n$$

These n vector equations can be summarized in the single vector equation,

$$X^t X \mathbf{a} = X^t \mathbf{y}$$

 $X^{t}X$ is an $n \times n$ matrix and since we have assumed that X has rank $n, X^{t}X$ will also have rank n. Since $X^{t}X$ is invertible, we have a critical point at

$$\mathbf{a}' = \left(X^t X\right)^{-1} X^t \mathbf{y}$$

Is this lone critical point really a minimum? The matrix of second partial derivatives is constant, and a positive multiple of $X^t X$. Theorem CPSM [1879] tells us that this matrix is positive semi-definite. In an advanced course on multivariable calculus, it is shown that

a minimum occurs exactly where the matrix of second partial derivatives is positive semidefinite. You may have seen this in the two-variable case, where a check on the positive semidefiniteness is disguised with a determinant of the 2×2 matrix of second partial derivatives.

Subsection EXC Exercises

T20 Theorem IP [1940] constructs a unique polynomial through a set of n + 1 points in the plane, $\{(x_i, y_i) | 1 \le i \le n + 1\}$, where the *x*-coordinates are all different. Prove that the expression below is the same polynomial and include an explanation of the necessity of the hypothesis that the *x*-coordinates are all different.

$$p(x) = \sum_{i=1}^{n+1} y_i \prod_{\substack{j=1\\j \neq i}}^{n+1} \frac{x - x_j}{x_i - x_j}$$

This is known as the Lagrange form of the interpolating polynomial. Contributed by Robert Beezer

Section SAS Sharing A Secret

This Section is a Draft, Subject to Changes

In this section we will see how to use solutions to systems of equations to share a secret among a group of people. We will be able to break a secret up into, say 10 pieces, so as to distribute the secret among 10 people. But rather than requiring all 10 people to collaborate on restoring the secret, we can design the split so that any smaller group, of say just 4 of these people, can collaborate and restore the secret. The numbers 10 and 4 here are arbitrary, we can choose them to be anything.

Suppose we have a secret, S. This could be the combination to a lock, a password on an account, or a recipe for chocolate chip cookies. If the secret is text, we will assume that the characters have been translated into integers (say with the ASCII code), and these numbers have been rolled up into one grand positive integer (perhaps by concatenating binary strings for the ASCII code numbers, and interpreting the longer string as one big base 2 integer). So we will assume S is some positive integer.

Suppose you wish to give parts of your secret to n people, and you wish to require that any group of m (or more) of these people should be able to combine their parts and recover the secret. Perhaps you are President and CEO of a small company and only you know the password that authorizes large transfers of money among the company's bank accounts. If you were to die or become incapacitated, it would perhaps hamper the company's ability to function if they couldn't quickly rearrange their assets, especially since they are also without a CEO. So you might wish to give this secret to six of your trusted Vice-Presidents. But you don't trust them that much and you certainly don't want any one of these people to be able to access the company's accounts all by themselves without anybody else in the company knowing about it. Simultaneously, you know that in an emergency, it might not be possible to get all six Vice-Presidents together and maybe even one or two of them have met the same unfortunate fate you did. So you would like any group of three Vice-Presidents to be able to combine their parts and recover S. So you would choose n = 6 and m = 3.

We will describe the split, with no motivation. The explanation of how the secret recovery is handled will explain our choices here. Choose a large prime number, p, bigger than any possible secret. For a single number in a combination lock, p could be small. For a one-page recipe, p would need to be huge. All of our subsequent arithmetic will be modulo p, so consult Subsection F.FF [1833] for a brief description of how we do linear algebra when our field is \mathbb{Z}_p . Build a polynomial, r(x), of degree m - 1 as follows. Set the constant term to S, and choose the other m - 1 coefficients at random from \mathbb{Z}_p . The quality of your random generator will ultimately affect the quality of how hidden your secret remains.

Compute the pairs (i, r(i)), $1 \leq i \leq n$. To person *i*, of the *n* persons you will give a part of your secret, present the pair (i, r(i)), and instruct them to keep this secret, for all $1 \leq i \leq n$. They could perhaps encrypt their pairs with AES (Advanced Encryption Standard) using a password known only to them individually. Or you could do this for each of them in advance and tell them the chose password orally, in private. At any rate, each person gets a pair of integers, an input to the polynomial, and the output of evaluating the polynomial, and they keep this information secret. They do not know the polynomial itself, and certainly not the constant term S, so the secret is still safe.

Now suppose that m of these people get together, in the event you are unable to act, or perhaps without your permission. Suppose they pool all of their pairs, or even just turn them over to one member of the group. What do they now know collectively? Suppose that

$$r(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_{m-1} x^{m-1}$$

where, of course, $a_0 = S$ is the secret. A single pair, (i, r(i)), results in a linear equation whose unknowns are the *m* coefficients of r(x). With *m* pairs revealed, we now have *m* equations in *m* variables. Furthermore, the coefficient matrix of this system is a Vandermonde matrix (Definition VM [1870]). With our inputs to the polynomial all different (we used 1, 2, 3, ..., *n*), the Vandermonde matrix is nonsingular (Theorem NVM [1877]). Thus by Theorem NMUS [180] there is a unique solution for the coefficients of r(x). We only desire the constant term — the other coefficients (the randomly chosen ones) are of no interest, they were used to mask the secret as it was split into parts.

A few practical considerations. If certain individuals in your group are more important, or more trustworthy, you can give them more than one part. You could split a secret into 30 parts, giving 5 Vice-Presidents each 4 parts and give 10 department heads each 1 part. Then you might require 12 parts to be present. This way three Vice-Presidents could recover the secret, or 4 department heads could stand-in for a Vice-President. Furthermore, the 10 department heads could not recover the secret without having at least one Vice-President present.

The inputs do not have to be consecutive integers, starting at 1. Any set of *different* integers will suffice. Why make it any easier for an attacker? Mix it up and choose the inputs randomly as well, just keep them different.

Why do all this arithmetic over \mathbb{Z}_p ? If we worked with polynomials having real number coefficients, properties of polynomials as continuous functions might give an attacker the ability to compute the secret with a reasonable amount of computing time. For example, the magnitude of the output is going to dominated by the term of r(x) having degree m-1. Suppose an attacker had a few of the pairs, but not a full set of m of them. Or even worse, suppose some group of fewer than m of your trusted acquaintances were to conspire against you. It might be possible to guess a limited range of values for the coefficient of the largest term. With a limited range of values here, the next term might fall to a similar analysis. And so on. However, modular arithmetic is in some ways very unpredictable looking and as high powers "wrap-around" this sort of analysis will be frustrated. And we know it is no harder to do linear algebra in \mathbb{Z}_p than in \mathbb{C} .

OK, here's a non-trivial example.

Example SS6W

Sharing a secret 6 ways

Let's return to the CEO and his six Vice-Presidents. Suppose the password for the company's accounts is a sequence of 5 two-digit numbers, which we will concatenate into a 10-digit number, in this case S = 0603725962. For a prime p we choose the 11-digit prime number p = 22801761379. From the requirement that m = 3 Vice-Presidents are needed to recover the secret, we need a second-degree polynomial and so need two more coefficients, which we

will construct at random between 1 and p. The resulting polynomial is

 $r(x) = 603725962 + 22561982919x + 8844088338x^2$

We will now build six pairs of inputs and outputs, where we will choose the inputs at random (not allowing duplicates) and we do all our arithmetic modulo p,

VP	x	r(x)
Finance	20220406046	7205699654
Human Resources	8862377358	17357568951
Marketing	13747127957	18503158079
Legal	15835120319	14060705999
Research	6530855859	5628836054
Manufacturing	9222703664	2608052019

The two numbers of each row of the table are then given to the indicated Vice-President. Done. The secret has been split six ways, and any three VP's can jointly recover the secret.

Let's test the recovery process, especially since it contains the relevant linear algebra. Suppose we write the unknown polynomial as $r(x) = a_0 + a_1x + a_2x^2$ and the VP's for Finance, Marketing and Legal all get together to recover the secret. The equations we arrive at are,

Finance

7205699654 = r(20220406046)

 $= a_0 + a_1(20220406046) + a_2(20220406046)^2$

 $= a_0 + 20220406046a_1 + 7793596215a_2$ Marketing 18503158079 = r(13747127957) $= a_0 + a_1(13747127957) + a_2(13747127957)^2$ $= a_0 + 13747127957a_1 + 18840301370a_2$ Legal 14060705999 = r(15835120319) $= a_0 + a_1(15835120319) + a_2(15835120319)^2$ $= a_0 + 15835120319a_1 + 8874412999a_2$

So they have a linear system, $\mathcal{LS}(A, \mathbf{b})$ with

	[1	20220406046	7793596215		7205699654
A =	1	13747127957	18840301370	$\mathbf{b} =$	18503158079
	1	15835120319	8874412999		14060705999

With a Vandermonde matrix as the coefficient matrix, they know there is a solution, and it is unique. By Theorem SNCM [543] (or through row-reducing the augmented matrix) they arrive at the solution,

	5716900879	9234437646	7850422855	7205699654		603725962]
$A^{-1}\mathbf{b} =$	20952200747	16452595922	8198726089	18503158079	=	22561982919
	17286943796	18018241597	10298337365	14060705999		8844088338

So the CEO's password is the secret $S = a_0 = 603725962 = 0603725962$ (as expected).

Index

A (appendix), 1637A (archetype), 1642 A (definition), 442 A (notation), 442A (part), 1940 AA (Property), 660 AA (subsection, section WILA), 5 AA (theorem), 443 AAC (Property), 205 AACN (Property), 1597 AAF (Property), 1831 AALC (example), 227 AAM (Property), 430 ABLC (example), 225 ABS (example), 271

AC (Property), 660 ACC (Property), 205 ACCN (Property), 1596 ACF (Property), 1830 ACM (Property), 430 ACN (example), 1595 additive associativity column vectors Property AAC, 205 complex numbers Property AACN, 1597 matrices Property AAM, 430 vectors Property AA, 660

additive closure column vectors Property ACC, 205 complex numbers Property ACCN, 1596 field Property ACF, 1830 matrices Property ACM, 430 vectors Property AC, 660 additive commutativity complex numbers Property CACN, 1597 additive inverse complex numbers Property AICN, 1597 from scalar multiplication theorem AISM, 678 additive inverses column vectors Property AIC, 206 matrices

Property AIM, 431 unique theorem AIU, 675 vectors Property AI, 660 adjoint definition A, 442 inner product theorem AIP, 485 notation, 442of a matrix sum theorem AMA, 442 of an adjoint theorem AA, 443 of matrix scalar multiplication theorem AMSM, 443 AHSAC (example), 147 AI (Property), 660 AIC (Property), 206 AICN (Property), 1597 AIF (Property), 1831 AIM (Property), 431 AIP (theorem), 485

AISM (theorem), 678 AIU (theorem), 675 AIVLT (example), 1214 ALT (example), 1074 ALTMM (example), 1296 AM (definition), 57 AM (example), 51 AM (notation), 57 AM (subsection, section MO), 442 AMA (theorem), 442 AMAA (example), 58 AME (definition), 968 AME (notation), 968 AMSM (theorem), 443 ANILT (example), 1216 ANM (example), 1428 AOS (example), 410Archetype A column space, 573linearly dependent columns, 329 singular matrix, 174 solving homogeneous system, 149 system as linear combination, 227

archetype A augmented matrix example AMAA, 58 Archetype B column space, 574inverse example CMIAB, 517 linearly independent columns, 329 nonsingular matrix, 174 not invertible example MWIAA, 506 solutions via inverse example SABMI, 503 solving homogeneous system, 148 system as linear combination, 225 vector equality, 200archetype B solutions example SAB, 79 Archetype C homogeneous system, 147 Archetype D column space, original columns, 571

solving homogeneous system, 150 vector form of solutions, 232 Archetype I column space from row operations, 587 null space, 153row space, 578 vector form of solutions, 248 Archetype I:casting out vectors, 370 Archetype L null space span, linearly independent, 336 vector form of solutions, 252 ASC (example), 1273 augmented matrix notation, 57AVR (example), 754 B (archetype), 1652

B (definition), 777
B (section), 777
B (subsection, section B), 777
basis

columns nonsingular matrix

example CABAK, 790

common size theorem BIS, 823 crazy vector apace example BC, 784 definition B, 777 matrices example BM, 780 example BSM22, 782 polynomials example BP, 780 example BPR, 854 example BSP4, 781 example SVP4, 856 subspace of matrices example BDM22, 854 BC (example), 784 BCS (theorem), 569 BDE (example), 1008 BDM22 (example), 854 best cities money magazine example MBC, 464 BIS (theorem), 823

- BM (example), 780 BNM (subsection, section B), 789 BNS (theorem), 334 BP (example), 780 BPR (example), 854 BRLT (example), 1190 BRS (theorem), 583 BS (theorem), 376 BSCV (subsection, section B), 785 BSM22 (example), 782 BSP4 (example), 781
- C (archetype), 1662 C (definition), 1606 C (notation), 1606 C (part), 2 C (Property), 660 C (technique, section PT), 1618 CABAK (example), 790 CACN (Property), 1597 CAEHW (example), 955 CAF (Property), 1830 canonical form

nilpotent linear transformation example CFNLT, 1468 theorem CFNLT, 1458 CAV (subsection, section O), 395 Cayley-Hamilton theorem CHT, 1563 CB (section), 1355 CB (theorem), 1359 CBCV (example), 1367 CBM (definition), 1359 CBM (subsection, section CB), 1358 CBP (example), 1361 CC (Property), 205 CCCV (definition), 396 CCCV (notation), 396 CCM (definition), 438 CCM (example), 439 CCM (notation), 438 CCM (theorem), 440 CCN (definition), 1598 CCN (notation), 1598 CCN (subsection, section CNO), 1598 CCRA (theorem), 1599

CCRM (theorem), 1599 CCS (subsection, section FS), 612 CCT (theorem), 1600 CD (subsection, section DM), 896 CD (technique, section PT), 1623 CEE (subsection, section EE), 960 CELT (example), 1397 CELT (subsection, section CB), 1385 CEMS6 (example), 974 CF (section), 1940 CFDVS (theorem), 1272 CFNLT (example), 1468 CFNLT (subsection, section NLT), 1458 CFNLT (theorem), 1458 CFV (example), 124 change of basis between polynomials example CBP, 1361 change-of-basis between column vectors example CBCV, 1367 matrix representation theorem MRCB, 1371

similarity theorem SCB, 1377 theorem CB, 1359 change-of-basis matrix definition CBM, 1359 inverse theorem ICBM, 1360 characteristic polynomial definition CP, 960 degree theorem DCP, 1013 size 3 matrix example CPMS3, 961 CHT (subsection, section JCF), 1563 CHT (theorem), 1563 CILT (subsection, section ILT), 1152 CILTI (theorem), 1152 CIM (subsection, section MISLE), 509 CINM (theorem), 516 CIVLT (example), 1224 CIVLT (theorem), 1227 CLI (theorem), 1275 CLTLT (theorem), 1111

CM (definition), 53 CM (Property), 430 CM32 (example), 1279 CMCN (Property), 1597 CMF (Property), 1831 CMI (example), 512 CMIAB (example), 517 CMVEI (theorem), 127 CN (appendix), 1565 CNA (definition), 1596 CNA (notation), 1596 CNA (subsection, section CNO), 1594 CNE (definition), 1596 CNE (notation), 1596 CNM (definition), 1596 CNM (notation), 1596 CNMB (theorem), 789 CNO (section), 1594 CNS1 (example), 154 CNS2 (example), 156CNSV (example), 404 COB (theorem), 793 coefficient matrix

definition CM, 53 nonsingular theorem SNCM, 543 column space as null space theorem FS, 625Archetype A example CSAA, 573 Archetype B example CSAB, 574 as null space example CSANS, 612 as null space, Archetype G example FSAG, 638 as row space theorem CSRST, 586 basis theorem BCS, 569 consistent system theorem CSCS, 563 consistent systems example CSMCS, 561 isomorphic to range, 1317

matrix, 560 nonsingular matrix theorem CSNM, 574 notation, 560original columns, Archetype D example CSOCD, 571 row operations, Archetype I example CSROI, 587 subspace theorem CSMS, 719 testing membership example MCSM, 564 two computations example CSTW, 568 column vector addition notation, 201 column vector scalar multiplication notation, 203commutativity column vectors Property CC, 205 matrices Property CM, 430

vectors Property C, 660 complex m-space example VSCV, 663 complex arithmetic example ACN, 1595 complex number conjugate example CSCN, 1599 modulus example MSCN, 1601 complex number conjugate definition CCN, 1598 modulus definition MCN, 1601 complex numbers addition definition CNA, 1596 notation, 1596 arithmetic properties theorem PCNA, 1596 equality

definition CNE, 1596 notation, 1596 multiplication definition CNM, 1596 notation, 1596 complex vector space dimension theorem DCM, 824 composition injective linear transformations theorem CILTI, 1152 surjective linear transformations theorem CSLTS, 1195 conjugate addition theorem CCRA, 1599 column vector definition CCCV, 396 matrix definition CCM, 438 notation, 438 multiplication theorem CCRM, 1599

notation, 1598 of conjugate of a matrix theorem CCM, 440 scalar multiplication theorem CRSM, 397 twice theorem CCT, 1600 vector addition theorem CRVA, 396 conjugate of a vector notation, 396 conjugation matrix addition theorem CRMA, 439 matrix scalar multiplication theorem CRMSM. 440 matrix transpose theorem MCT, 441 consistent linear system, 120 consistent linear systems theorem CSRN, 123 consistent system definition CS, 113

constructive proofs technique C, 1618 contradiction technique CD, 1623 contrapositive technique CP, 1621 converse technique CV, 1622coordinates orthonormal basis theorem COB, 793 coordinatization linear combination of matrices example CM32, 1279 linear independence theorem CLI, 1275 orthonormal basis example CROB3, 796 example CROB4, 794 spanning sets theorem CSS, 1276 coordinatization principle, 1279 coordinatizing

polynomials example CP2, 1277 COV (example), 370 COV (subsection, section LDS), 370 CP (definition), 960 CP (subsection, section VR), 1275 CP (technique, section PT), 1621 CP2 (example), 1277 CPMS3 (example), 961 CPSM (theorem), 1879 crazy vector space example CVSR, 1273 properties example PCVS, 679 CRMA (theorem), 439 CRMSM (theorem), 440 CRN (theorem), 830 CROB3 (example), 796 CROB4 (example), 794 CRS (section), 560CRSM (theorem), 397 CRVA (theorem), 396 CS (definition), 113

CS (example), 1606 CS (subsection, section TSS), 112 CSAA (example), 573 CSAB (example), 574 CSANS (example), 612 CSCN (example), 1599 CSCS (theorem), 563 CSIP (example), 398 CSLT (subsection, section SLT), 1194 CSLTS (theorem), 1195 CSM (definition), 560 CSM (notation), 560 CSMCS (example), 561 CSMS (theorem), 719 CSNM (subsection, section CRS), 572 CSNM (theorem), 574 CSOCD (example), 571 CSRN (theorem), 123 CSROI (example), 587 CSRST (diagram), 642 CSRST (theorem), 586 CSS (theorem), 1276 CSSE (subsection, section CRS), 561

CSSOC (subsection, section CRS), 567 CSTW (example), 568 CTD (subsection, section TD), 1907 CTLT (example), 1112 CUMOS (theorem), 547 curve fitting polynomial through 5 points example PTFP, 1941 CV (definition), 52 CV (notation), 52 CV (technique, section PT), 1622 CVA (definition), 201 CVA (notation), 201 CVC (notation), 52 CVE (definition), 199 CVE (notation), 199 CVS (example), 670 CVS (subsection, section VR), 1272 CVSM (definition), 203 CVSM (example), 204 CVSM (notation), 203 CVSR (example), 1273

D (archetype), 1670	AM, 57
D (chapter), 882	AME, 968
D (definition), 815	B, 777
D (notation), 815	C, 1606
D (section), 815	CBM, 1359
D (subsection, section D), 815	CCCV, 396
D (subsection, section SD), 1037	CCM, 438
D (technique, section PT), 1611	CCN, 1598
D33M (example), 895	CM, 53
DAB (example), 1037	CNA, 1596
DC (example), 827	CNE, 1596
DC (technique, section PT), 1628	CNM, 1596
DC (theorem), 1039	CP, 960
DCM (theorem), 824	CS, 113
DCN (Property), 1597	CSM, 560
DCP (theorem), 1013	CV, 52
DD (subsection, section DM), 893	CVA, 201
DEC (theorem), 901	CVE, 199
decomposition	CVSM, 203
technique DC, 1628	D, 815
DED (theorem), 1048	DIM, 1037
definition	DM, 894
A, 442	DS, 864

DZM, 1037	
EEF, 619	
EELT, 1356	
EEM, 945	
ELEM, 883	
EM, <mark>963</mark>	
EO, 23	
ES, 1603	
ESYS, 22	
F, 1830	
GES, 1484	
GEV, 1484	
GME, <mark>969</mark>	
HI, 1860	
HID, 1859	
HM, 486	
HP, 1857	
HS, 146	
IDLT, 1213	
IDV, 118	
IE, 1508	
ILT, 1128	
IM, 175	

IMP, 1833 IP, **398** IS, 1476 IVLT, 1214 IVS, 1230 JB, 1441 JCF, 1532 KLT, 1138 LC, 707 LCCV, 221 LI, 735 LICV, 317 LNS, 609 LSS, 1944 LT, 1072 LTA, 1105 LTC, 1111 LTM, 1415 LTR, 1495 LTSM, 1108 M, 50 MA, 427 MCN, 1601

ME, 427
MI, 506
MM, 468
MR, 1286
MRLS, 56
MSM, 428
MVP, 461
NLT, 1437
NM, 173
NOLT, 1234
NOM, 829
NRML, 1427
NSM, 152
NV, 404
ONS, 418
OSV, 409
OV, 408
PI, 1101
PSM, 1879
REM, 60
RLD, 735
RLDCV, 316
RLT, 1178

RO, 59
ROLT, 1233
ROM, 829
RR, 85
RREF, 63
RSM, 577
S, 694
SC, 1608
SE, 1604
SET, 1602
SI, 1608
SIM, 1029
SLE, 18
SLT, 1169
SM, 893
SOLV, 55
SQM, 173
SRM, 1934
SS, 709
SSCV, 270
SSEV, 270 SSET, 1603
SSL1, 1005 SSLE, 18
$\begin{array}{c} \text{SSLE, } 18 \\ \text{SSSLE, } 19 \end{array}$
SSSLE, 19

SU, 1607 SUV, 409 SV, 1923 SYM, 434 T, 1845 technique D, 1611 TM, 433 TS, 704 TSHSE, 148 TSVS, 746 UM, 544 UTM. 1415 VM, 1870 VOC, 54 VR, 1261 VS, 659 VSCV. 198 VSM, 426 ZCV, 53 ZM, 433 DEHD (example), 1049 DEM (theorem), 932 DEMMM (theorem), 933 DEMS5 (example), 979 DER (theorem), 897 DERC (theorem), 923 determinant computed two ways example TCSD, 902 definition DM, 894 equal rows or columns theorem DERC, 923 expansion, columns theorem DEC, 901 expansion, rows theorem DER, 897 identity matrix theorem DIM, 930matrix multiplication theorem DRMM, 938 nonsingular matrix, 935 notation, 894 row or column multiple theorem DRCM, 922 row or column swap theorem DRCS, 920

size 2 matrix theorem DMST, 896 size 3 matrix example D33M, 895 transpose theorem DT, 899 via row operations example DRO, 926 zero theorem SMZD, 935 zero row or column theorem DZRC, 919 zero versus nonzero example ZNDAB, 936 determinant, upper triangular matrix example DUTM, 903 determinants elementary matrices theorem DEMMM, 933 DF (Property), 1831 DF (subsection, section CF), 1942 DFS (subsection, section PD), 861 DFS (theorem), 861

DGES (theorem), 1531diagonal matrix definition DIM, 1037 diagonalizable definition DZM, 1037 distinct eigenvalues example DEHD, 1049 theorem DED, 1048full eigenspaces theorem DMFE, 1044 not example NDMS4, 1047 diagonalizable matrix high power example HPDM, 1051 diagonalization Archetype B example DAB, 1037 criteria theorem DC, 1039example DMS3, 1042 diagram CSRST, **642**

DLTA, 1073 DLTM, 1074 DTSLS. 126 FTMR, 1294 FTMRA, 1295 GLT, 1081 ILT, 1134 MRCLT, 1309 NILT, 1130 DIM (definition), 1037 DIM (theorem), 930 dimension crazy vector space example DC, 827 definition D, 815 notation, 815 polynomial subspace example DSP4, 827 proper subspaces theorem PSSD, 857 subspace example DSM22, 825 direct sum

decomposing zero vector theorem DSZV, 868 definition DS, 864 dimension theorem DSD, 872 example SDS, 865 from a basis theorem DSFB, 865 from one subspace theorem DSFOS, 867 notation, 864 zero intersection theorem DSZI, 869 direct sums linear independence theorem DSLI, 870 repeated theorem RDS, 873 distributivity complex numbers Property DCN, 1597 field Property DF, 1831

distributivity, matrix addition matrices Property DMAM, 431 distributivity, scalar addition column vectors Property DSAC, 206 matrices Property DSAM, 431 vectors Property DSA, 660 distributivity, vector addition column vectors Property DVAC, 206 vectors Property DVA, 660 DLDS (theorem), 365 DLTA (diagram), 1073 DLTM (diagram), 1074 DM (definition), 894 DM (notation), 894 DM (section), 883 DM (theorem), 824DMAM (Property), 431

DMFE (theorem), 1044DMHP (subsection, section HP), 1863 DMHP (theorem), 1863 DMMP (theorem), 1865 DMS3 (example), 1042 DMST (theorem), 896 DNLT (theorem), 1449 DNMMM (subsection, section PDM), 934 DP (theorem), 824DRCM (theorem), 922 DRCMA (theorem), 924 DRCS (theorem), 920 DRMM (theorem), 938 DRO (example), 926 DRO (subsection, section PDM), 919 DROEM (subsection, section PDM), 930 DS (definition), 864 DS (notation), 864 DS (subsection, section PD), 863 DSA (Property), 660 DSAC (Property), 206 DSAM (Property), 431 DSD (theorem), 872

DSFB (theorem), 865 DSFOS (theorem), 867 DSLI (theorem), 870 DSM22 (example), 825 DSP4 (example), 827 DSZI (theorem), 869 DSZV (theorem), 868 DT (theorem), 899 DTSLS (diagram), 126 DUTM (example), 903 DVA (Property), 660 DVAC (Property), 206 DVM (theorem), 1871 DVS (subsection, section D), 824DZM (definition), 1037 DZRC (theorem), 919

E (archetype), 1678 E (chapter), 944 E (technique, section PT), 1619 E.SAGE (computation, section SAGE), 1591 ECEE (subsection, section EE), 968 EDELI (theorem), 1001 EDYES (theorem), 858 EE (section), 945 EEE (subsection, section EE), 952 EEF (definition), 619 EEF (subsection, section FS), 618EELT (definition), 1356 EELT (subsection, section CB), 1355 EEM (definition), 945 EEM (subsection, section EE), 945 EEMAP (theorem), 1915 EENS (example), 1036 EER (theorem), 1384 EESR (theorem), 1930 EHM (subsection, section PEE), 1020 eigenspace as null space theorem EMNS, 965 definition EM, 963 invariant subspace theorem EIS, 1479 subspace theorem EMS, 963 eigenspaces

sage, 1591 eigenvalue algebraic multiplicity definition AME, 968 notation, 968 complex example CEMS6, 974 definition EEM, 945 existence example CAEHW, 955 theorem EMHE, 953 geometric multiplicity definition GME, 969 notation, 969 index, 1508 linear transformation definition EELT, 1356 multiplicities example EMMS4, 969 power theorem EOMP, 1006 root of characteristic polynomial theorem EMRCP, 962

scalar multiple theorem ESMM, 1005 symmetric matrix example ESMS4, 970 zero theorem SMZE, 1003 eigenvalues building desired example BDE, 1008 complex, of a linear transformation example CELT, 1397 conjugate pairs theorem ERMCP, 1012 distinct example DEMS5, 979 example SEE, 946 Hermitian matrices theorem HMRE, 1020 inverse theorem EIM, 1009 maximum number theorem MNEM, 1019 multiplicities

example HMEM5, 973 theorem ME, 1016number theorem NEM, 1015of a polynomial theorem EPM, 1007size 3 matrix example EMS3, 962 example ESMS3, 966 transpose theorem ETM, 1011eigenvalues, eigenvectors vector, matrix representations theorem EER, 1384eigenvector, 945 linear transformation, 1356 eigenvectors, 946 conjugate pairs, 1012 Hermitian matrices theorem HMOE, 1022 linear transformation example ELTBM, 1356 example ELTBP, 1357

linearly independent theorem EDELI, 1001 of a linear transformation example ELTT, 1386 EILT (subsection, section ILT), 1129 EIM (theorem), 1009 EIS (example), 1480 EIS (theorem), 1479 ELEM (definition), 883 ELEM (notation), 885 elementary matrices definition ELEM, 883 determinants theorem DEM, 932 nonsingular theorem EMN, 891 notation, 885 row operations example EMRO, 885 theorem EMDRO, 886 ELIS (theorem), 850 ELTBM (example), 1356 ELTBP (example), 1357

ELTT (example), 1386 EM (definition), 963 EM (subsection, section DM), 883 EMDRO (theorem), 886 EMHE (theorem), 953 EMMS4 (example), 969 EMMVP (theorem), 466 EMN (theorem), 891 EMNS (theorem), 965 EMP (theorem), 471empty set, 1603notation, 1603 EMRCP (theorem), 962 EMRO (example), 885 EMS (theorem), 963 EMS3 (example), 962 ENLT (theorem), 1448 EO (definition), 23 EOMP (theorem), 1006 EOPSS (theorem), 24 EPM (theorem), 1007 EPSM (theorem), 1880 equal matrices

via equal matrix-vector products theorem EMMVP, 466 equation operations definition EO, 23 theorem EOPSS, 24 equivalence statements technique E, 1619 equivalences technique ME, 1626 equivalent systems definition ESYS, 22 ERMCP (theorem), 1012 ES (definition), 1603 ES (notation), 1603 ESEO (subsection, section SSLE), 22 ESLT (subsection, section SLT), 1170 ESMM (theorem), 1005 ESMS3 (example), 966 ESMS4 (example), 970 ESYS (definition), 22 ETM (theorem), 1011 EVS (subsection, section VS), 663 example

AALC, 227 ABLC, 225 ABS, 271 ACN, 1595 AHSAC, 147 AIVLT, 1214 ALT, 1074 ALTMM, 1296 AM, 51 AMAA, 58 ANILT, 1216 ANM, 1428 AOS, **410** ASC, 1273 AVR, 754 BC, 784 BDE, 1008 BDM22, 854 BM, 780 BP, 780 BPR, 854 BRLT, 1190 BSM22, 782

BSP4, 781 CABAK, 790 CAEHW, 955 CBCV, 1367 CBP, **1361** CCM, 439 CELT, 1397 CEMS6, 974 CFNLT, 1468 CFV, 124 CIVLT, 1224 CM32, 1279 CMI, 512 CMIAB, 517 CNS1, 154 CNS2, 156 CNSV, 404 COV, **370** CP2, 1277 CPMS3, 961 CROB3, 796 CROB4, 794 CS, 1606

CSAA, 573 CSAB, 574 CSANS, 612 CSCN, 1599 CSIP, **398** CSMCS, 561 CSOCD, 571 CSROI, **587** CSTW, 568 CTLT, 1112 CVS, 670 CVSM, 204 CVSR, 1273 D33M, 895 DAB, 1037 DC, 827 DEHD, 1049 DEMS5, 979 DMS3, 1042 DRO, 926 DSM22, 825 DSP4, 827 DUTM, 903

EENS, 1036 EIS, 1480 ELTBM, 1356 ELTBP, 1357 ELTT, 1386 EMMS4, 969 EMRO, 885 EMS3, 962 ESMS3, 966 ESMS4, 970 FDV, 118 FF8, 1837 FRAN, 1182 FS1, 634 FS2, 635 FSAG, 638 FSCF, 1054 GE4, 1488 GE6, 1490 GENR6, 1508 GSTV, 417 HISAA, 149 HISAD, 150

HMEM5, 973 HP, 1858 HPDM, 1051 HUSAB, 148 IAP, 1147 IAR, 1131 IAS, 585 IAV, 1135 ILTVR, 1324 IM, 175 IM11, 1833 IS, **31** ISJB, 1482 ISMR4, 1501 ISMR6, 1503 ISSI, 115 IVSAV, 1230 JB4, 1441 JCF10, 1535 KPNLT, 1454 KVMR, 1313 LCM, 707 LDCAA, 328

LDHS, 323 LDP4, 822 LDRN, 326 LDS, 317 LIC, 743 LICAB, 329 LIHS, 322 LIM32, 739 LINSB, **332** LIP4, 736 LIS, 320 LLDS, 327 LNS, 610 LTDB1, 1097 LTDB2, 1098 LTDB3, 1100 LTM, 1084 LTPM, 1078 LTPP, 1079 LTRGE, 1495 MA, 428 MBC, 464 MCSM, 564

MFLT, 1087 MI, 507 MIVS, 1274 MMNC, 470 MNSLE, 463 MOLT, 1091 MPMR, 1304 MRBE, 1378 MRCM, 1372 MSCN, 1601 MSM, 429 MTV, 461 MWIAA, 506 NDMS4, 1047 NIAO, 1147 NIAQ, 1129 NIAQR, 1145 NIDAU, 1151 NJB5, 1442 NKAO, 1138 NLT, 1077 NM, 174 NM62, 1439

NM64, 1437 NM83, 1444 NRREF, 65 NSAO, 1187 NSAQ, 1170 NSAQR, 1185 NSC2A, 703 NSC2S, 703 NSC2Z, 702 NSDAT, 1193 NSDS, 287 NSE, 19 NSEAI, 153 NSLE, 56 NSLIL, 336 NSNM, 179 NSR, 177 NSS, 178 OLTTR, 1286 ONFV, 419 ONTV, 418 OSGMD, 127 OSMC, 548

PCVS, 679 PM, 949 PSHS, 258 PTFP, 1941 PTM, 468 PTMEE, 472 RAO, 1178 RES, 382 RNM, 829 RNSM, 832 ROD2, 1891 ROD4, 1893 RREF, 65 RREFN, 113 RRTI, 860 RS, 787 RSAI, 577 RSB, 785 RSC4, 380 RSC5, 366 RSNS, 706 **RSREM**, **582** RVMR, 1319

S, 174 SAA, 81 SAB, 79 SABMI, 503 SAE, 83 SAN, 1188 SAR, 1172 SAV, 1175 SC, 1609 SC3, 694 SCAA, 275 SCAB, 279 SCAD, 290 SDS, 865 SEE, 946 SEEF, 619 SETM, 1602 SI, 1608 SM2Z7, 1836 SM32, 714 SMLT, 1109 SMS3, 1031 SMS5, 1030

SP4, 700
SPIAS, 1101
SRR, 176
SS, 893
SS6W, 1952
SSC, 751
SSET, 1603
SSM22, 749
SSNS, 285
SSP, 712
SSP4, 746
STLT, 1107
STNE, 16
SU, 1607
SUVOS, 410
SVP4, 856
SYM, 435
TCSD, 902
TD4, 1902
TDEE6, 1911
TDSSE, 1905
TIS, 1476
TIVS, 1273

TKAP, 1141
TLC, 222
TM, 433
TMP, 6
TOV, 408
TREM, 60
TTS, 21
UM3, 545
UPM, 545
US, 29
USR, <u>62</u>
VA, 202
VESE, 200
VFS, 237
VFSAD, 232
VFSAI, 248
VFSAL, 252
VM4, 1870
VRC4, 1263
VRP2, 1267
VSCV, 663
VSF, 667
VSIM5, 1834

VSIS, 667	EXC (subsection, section LISS), 759
VSM, 663	EXC (subsection, section LT), 1115
VSP, 665	EXC (subsection, section MINM), 553
VSPUD, 828	EXC (subsection, section MISLE), 525
VSS, 668	EXC (subsection, section MM), 490
ZNDAB, <u>936</u>	EXC (subsection, section MO), 446
EXC (subsection, section B), 802	EXC (subsection, section MR), 1331
EXC (subsection, section CB), 1404	EXC (subsection, section NM), 184
EXC (subsection, section CF), 1948	EXC (subsection, section O), 422
EXC (subsection, section CRS), 591	EXC (subsection, section PD), 876
EXC (subsection, section D), 837	EXC (subsection, section PDM), 941
EXC (subsection, section DM), 907	EXC (subsection, section PEE), 1024
EXC (subsection, section EE), 984	EXC (subsection, section PSM), 1884
EXC (subsection, section F), 1841	EXC (subsection, section RREF), 88
EXC (subsection, section FS), 645	EXC (subsection, section S), 722
EXC (subsection, section HP), 1868	EXC (subsection, section SD), 1060
EXC (subsection, section HSE), 158	EXC (subsection, section SLT), 1197
EXC (subsection, section ILT), 1154	EXC (subsection, section SS), 295
EXC (subsection, section IS), 1516	EXC (subsection, section SSLE), 36
EXC (subsection, section IVLT), 1245	EXC (subsection, section T), 1853
EXC (subsection, section LC), 263	EXC (subsection, section TSS), 131
EXC (subsection, section LDS), 386	EXC (subsection, section VO), 209
EXC (subsection, section LI), 339	EXC (subsection, section VR), 1282

EXC (subsection, section VS), 683
EXC (subsection, section WILA), 14
extended echelon form

submatrices
example SEEF, 619

extended reduced row-echelon form

properties
theorem PEEF, 621

F (archetype), 1686
F (definition), 1830
F (section), 1829
E (subsection continue E), 1820

F (definition), 1050
F (section), 1829
F (subsection, section F), 1830
FDV (example), 118
FF (subsection, section F), 1833
FF8 (example), 1837
Fibonacci sequence

example FSCF, 1054

field

definition F, 1830

FIMP (theorem), 1833
finite field

size 8

example FF8, 1837 four subsets example FS1, 634 example FS2, 635 four subspaces dimension theorem DFS, 861 FRAN (example), 1182 free variables example CFV, 124 free variables, number theorem FVCS, 124 free, independent variables example FDV, 118 FS (section), 609 FS (subsection, section FS), 624 FS (subsection, section SD), 1054 FS (theorem), 625 FS1 (example), 634FS2 (example), 635FSAG (example), 638 FSCF (example), 1054 FTMR (diagram), 1294

FTMR (theorem), 1293FTMRA (diagram), 1295FV (subsection, section TSS), 124FVCS (theorem), 124

G (archetype), 1697 G (theorem), 852 GE4 (example), 1488 GE6 (example), 1490 GEE (subsection, section IS), 1484 GEK (theorem), 1487 generalized eigenspace as kernel theorem GEK, 1487 definition GES, 1484 dimension theorem DGES, 1531 dimension 4 domain example GE4, 1488 dimension 6 domain example GE6, 1490 invariant subspace theorem GESIS, 1485

nilpotent restriction theorem RGEN, 1507 nilpotent restrictions, dimension 6 domain example GENR6, 1508 notation, 1485 generalized eigenspace decomposition theorem GESD, 1518 generalized eigenvector definition GEV, 1484 GENR6 (example), 1508 GES (definition), 1484 GES (notation), 1485 GESD (subsection, section JCF), 1518 GESD (theorem), 1518GESIS (theorem), 1485 GEV (definition), 1484 GFDL (appendix), 1814 GLT (diagram), 1081 GME (definition), 969 GME (notation), 969 goldilocks theorem G, 852 Gram-Schmidt

column vectors theorem GSP, 413 three vectors example GSTV, 417 gram-schmidt mathematica, 1573 GS (technique, section PT), 1617 GSP (subsection, section PT), 1617 GSP (subsection, section O), 413 GSP (theorem), 413 GSP.MMA (computation, section MMA), 1573 GSTV (example), 417 GT (subsection, section PD), 850

H (archetype), 1705
Hadamard Identity notation, 1859
Hadamard identity definition HID, 1859
Hadamard Inverse notation, 1860
Hadamard inverse definition HI, 1860
Hadamard Product

Diagonalizable Matrices theorem DMHP, 1863 notation, 1857 Hadamard product commutativity theorem HPC, 1858 definition HP, 1857 diagonal matrices theorem DMMP, 1865 distributivity theorem HPDAA, 1861 example HP, 1858 identity theorem HPHID, 1859 inverse theorem HPHI, 1860 scalar matrix multiplication theorem HPSMM, 1862 hermitian definition HM, 486 Hermitian matrix inner product theorem HMIP, 486

HI (definition), 1860 HI (notation), 1860 HID (definition), 1859 HID (notation), 1859 HISAA (example), 149 HISAD (example), 150 HM (definition), 486 HM (subsection, section MM), 485 HMEM5 (example), 973 HMIP (theorem), 486 HMOE (theorem), 1022 HMRE (theorem), 1020 HMVEI (theorem), 151 homogeneous system Archetype C example AHSAC, 147 consistent theorem HSC, 147 definition HS, 146 infinitely many solutions theorem HMVEI, 151 homogeneous systems linear independence, 321

HP (definition), 1857 HP (example), 1858 HP (notation), 1857 HP (section), 1857 HPC (theorem), 1858 HPDAA (theorem), 1861 HPDM (example), 1051 HPHI (theorem), 1860 HPHID (theorem), 1859 HPSMM (theorem), 1862 HS (definition), 146 HSC (theorem), 147 HSE (section), 146 HUSAB (example), 148

I (archetype), 1714 I (technique, section PT), 1630 IAP (example), 1147 IAR (example), 1131 IAS (example), 585 IAV (example), 1135 ICBM (theorem), 1360 ICLT (theorem), 1227

identities technique PI, 1627 identity matrix determinant, 930 example IM, 175 notation, 175 IDLT (definition), 1213 IDV (definition), 118 IE (definition), 1508 IE (notation), 1508 IFDVS (theorem), 1274 IILT (theorem), 1220 ILT (definition), 1128 ILT (diagram), 1134 ILT (section), 1128 ILTB (theorem), 1149 ILTD (subsection, section ILT), 1150 ILTD (theorem), 1150 ILTIS (theorem), 1221 ILTLI (subsection, section ILT), 1147 ILTLI (theorem), 1148 ILTLT (theorem), 1219 ILTVR (example), 1324

IM (definition), 175 IM (example), 175 IM (notation), 175 IM (subsection, section MISLE), 506 IM11 (example), 1833 IMILT (theorem), 1327 IMP (definition), 1833 IMR (theorem), 1322 inconsistent linear systems theorem ISRN, 122 independent, dependent variables definition IDV, 118 indesxstring example SM2Z7, 1836 example SSET, 1603 index eigenvalue definition IE, 1508 notation. 1508indexstring theorem DRCMA, 924 theorem OBUTR, 1424 theorem UMCOB, 798

induction technique I, 1630 infinite solution set example ISSI, 115 infinite solutions, 3×4 example IS, 31 injective example IAP, 1147 example IAR, 1131 not example NIAO, 1147 example NIAQ, 1129 example NIAQR, 1145 not, by dimension example NIDAU, 1151 polynomials to matrices example IAV, 1135 injective linear transformation bases theorem ILTB, 1149 injective linear transformations dimension theorem ILTD, 1150

inner product anti-commutative theorem IPAC, 402 example CSIP, 398 norm theorem IPN, 406 notation, 398 positive theorem PIP, 406 scalar multiplication theorem IPSM, 401 vector addition theorem IPVA, 400 integers $\mod p$ definition IMP, 1833 $\mod p$, field theorem FIMP, 1833 mod 11 example IM11, 1833 interpolating polynomial theorem IP, 1940 invariant subspace

definition IS, 1476 eigenspace, 1479 eigenspaces example EIS, 1480 example TIS, 1476 Jordan block example ISJB, 1482 kernels of powers theorem KPIS, 1481 inverse composition of linear transformations theorem ICLT, 1227 example CMI, 512example MI, 507 notation, 506 of a matrix, 506 invertible linear transformation defined by invertible matrix theorem IMILT, 1327 invertible linear transformations composition theorem CIVLT, 1227 computing

example CIVLT, 1224 IP (definition), 398 IP (notation), 398 IP (subsection, section O), 398 IP (theorem), 1940 IPAC (theorem), 402 IPN (theorem), 406IPSM (theorem), 401 IPVA (theorem), 400 IS (definition), 1476 IS (example), 31 IS (section), 1475 IS (subsection, section IS), 1475 ISJB (example), 1482 ISMR4 (example), 1501ISMR6 (example), 1503 isomorphic multiple vector spaces example MIVS, 1274 vector spaces example IVSAV, 1230 isomorphic vector spaces dimension

theorem IVSED, 1232 example TIVS, 1273 ISRN (theorem), 122 ISSI (example), 115 ITMT (theorem), 1417 IV (subsection, section IVLT), 1221 IVLT (definition), 1214 IVLT (section), 1213 IVLT (subsection, section IVLT), 1213 IVLT (subsection, section MR), 1321 IVS (definition), 1230 IVSAV (example), 1230 IVSED (theorem), 1232

J (archetype), 1723 JB (definition), 1441 JB (notation), 1441 JB4 (example), 1441 JCF (definition), 1532 JCF (section), 1517 JCF (subsection, section JCF), 1532 JCF10 (example), 1535 JCFLT (theorem), 1533 Jordan block definition JB, 1441 nilpotent theorem NJB, 1447 notation, 1441 size 4 example JB4, 1441 Jordan canonical form definition JCF, 1532 size 10 example JCF10, 1535

K (archetype), 1734 kernel injective linear transformation theorem KILT, 1144 isomorphic to null space theorem KNSI, 1311 linear transformation example NKAO, 1138 notation, 1138 of a linear transformation definition KLT, 1138

pre-image, 1143 subspace theorem KLTS, 1140 trivial example TKAP, 1141 via matrix representation example KVMR, 1313 KILT (theorem), 1144 KLT (definition), 1138 KLT (notation), 1138 KLT (subsection, section ILT), 1137 KLTS (theorem), 1140 KNSI (theorem), 1311 KPI (theorem), 1143 KPIS (theorem), 1481 KPLT (theorem), 1450 KPNLT (example), 1454 KPNLT (theorem), 1453 KVMR (example), 1313

L (archetype), 1742 L (technique, section PT), 1614 LA (subsection, section WILA), 3 LC (definition), 707 LC (section), 221LC (subsection, section LC), 221 LC (technique, section PT), 1634 LCCV (definition), 221 LCM (example), 707 LDCAA (example), 328 LDHS (example), 323 LDP4 (example), 822 LDRN (example), 326 LDS (example), 317 LDS (section), 364LDSS (subsection, section LDS), 364 least squares minimizes residuals theorem LSMR, 1944 least squares solution definition LSS, 1944 left null space as row space, 625definition LNS, 609 example LNS, 610notation, 609

subspace theorem LNSMS, 720 lemma technique LC, 1634 LI (definition), 735 LI (section), 316 LI (subsection, section LISS), 735 LIC (example), 743 LICAB (example), 329 LICV (definition), 317 LIHS (example), 322 LIM32 (example), 739 linear combination system of equations example ABLC, 225 definition LC, 707 definition LCCV, 221 example TLC, 222 linear transformation, 1093 matrices example LCM, 707 system of equations example AALC, 227

linear combinations solutions to linear systems theorem SLSLC, 229 linear dependence more vectors than size theorem MVSLD, 328 linear independence definition LI, 735 definition LICV, 317 homogeneous systems theorem LIVHS, 321 injective linear transformation theorem ILTLI, 1148 matrices example LIM32, 739 orthogonal, 412 r and n theorem LIVRN, 325 linear solve mathematica, 1567 sage, 1588 linear system consistent

theorem RCLS, 120 matrix representation definition MRLS, 56 notation, 56 linear systems notation example MNSLE, 463 example NSLE, 56 linear transformation polynomials to polynomials example LTPP, 1079 addition definition LTA, 1105 theorem MLTLT, 1108 theorem SLTLT, 1106 as matrix multiplication example ALTMM, 1296 basis of range example BRLT, 1190 checking example ALT, 1074 composition definition LTC, 1111

theorem CLTLT, 1111 defined by a matrix example LTM, 1084 defined on a basis example LTDB1, 1097 example LTDB2, 1098 example LTDB3, 1100 theorem LTDB, 1094 definition LT, 1072 identity definition IDLT, 1213 injection definition ILT, 1128 inverse theorem ILTLT, 1219 inverse of inverse theorem IILT, 1220 invertible definition IVLT, 1214 example AIVLT, 1214 invertible, injective and surjective theorem ILTIS, 1221 Jordan canonical form

theorem JCFLT, 1533 kernels of powers theorem KPLT, 1450 linear combination theorem LTLC, 1093 matrix of, 1089 example MFLT, 1087 example MOLT, 1091 not example NLT, 1077 not invertible example ANILT, 1216 notation, 1073 polynomials to matrices example LTPM, 1078 rank plus nullity theorem RPNDD, 1235 restriction definition LTR, 1495 notation, 1495 scalar multiple example SMLT, 1109 scalar multiplication

definition LTSM, 1108 spanning range theorem SSRLT, 1188 sum example STLT, 1107 surjection definition SLT, 1169 vector space of, 1110 zero vector theorem LTTZZ, 1080 linear transformation inverse via matrix representation example ILTVR, 1324 linear transformation restriction on generalized eigenspace example LTRGE, 1495 linear transformations compositions example CTLT, 1112 from matrices theorem MBLT, 1086 linearly dependent r < n

example LDRN, 326 via homogeneous system example LDHS, 323 linearly dependent columns Archetype A example LDCAA, 328 linearly dependent set example LDS, 317 linear combinations within theorem DLDS, 365 polynomials example LDP4, 822 linearly independent crazy vector space example LIC, 743 extending sets theorem ELIS, 850 polynomials example LIP4, 736 via homogeneous system example LIHS, 322 linearly independent columns Archetype B

example LICAB, 329 linearly independent set example LIS, 320 example LLDS, 327 LINM (subsection, section LI), 328 LINSB (example), 332 LIP4 (example), 736 LIS (example), 320 LISS (section), 734 LISV (subsection, section LI), 316 LIVHS (theorem), 321 LIVRN (theorem), 325 LLDS (example), 327 LNS (definition), 609 LNS (example), 610 LNS (notation), 609 LNS (subsection, section FS), 609 LNSMS (theorem), 720 lower triangular matrix definition LTM, 1415 LS.MMA (computation, section MMA), 1567 LS.SAGE (computation, section SAGE), 1588 LSMR (theorem), 1944

LSS (definition), 1944 LT (chapter), 1071 LT (definition), 1072 LT (notation), 1073 LT (section), 1072LT (subsection, section LT), 1072 LTA (definition), 1105 LTC (definition), 1111 LTC (subsection, section LT), 1081 LTDB (theorem), 1094 LTDB1 (example), 1097 LTDB2 (example), 1098 LTDB3 (example), 1100 LTLC (subsection, section LT), 1092 LTLC (theorem), 1093 LTM (definition), 1415 LTM (example), 1084 LTPM (example), 1078 LTPP (example), 1079 LTR (definition), 1495 LTR (notation), 1495 LTRGE (example), 1495 LTSM (definition), 1108

LTTZZ (theorem), 1080 M (archetype), 1750 M (chapter), 425M (definition), 50 M (notation), 51 MA (definition), 427 MA (example), 428MA (notation), 427MACN (Property), 1597 MAF (Property), 1831 MAP (subsection, section SVD), 1915 mathematica gram-schmidt (computation), 1573 linear solve (computation), 1567 matrix entry (computation), 1566 matrix inverse (computation), 1576 matrix multiplication (computation), 1575 null space (computation), 1569 row reduce (computation), 1567 transpose of a matrix (computation), 1575 vector form of solutions (computation), 1571

vector linear combinations (computation), 1568 mathematical language technique L, 1614 matrix addition definition MA, 427 notation, 427augmented definition AM, 57 column space definition CSM, 560 complex conjugate example CCM, 439 definition M, 50 equality definition ME, 427notation, 427example AM, 51 identity definition IM, 175 inverse definition MI, 506

nonsingular definition NM, 173 notation, 51of a linear transformation theorem MLTCV, 1089 product example PTM, 468 example PTMEE, 472 product with vector definition MVP, 461 rectangular, 173 row space definition RSM, 577 scalar multiplication definition MSM, 428 notation, 428singular, 173 square definition SQM, 173 submatrices example SS, 893 submatrix definition SM, 893

symmetric definition SYM, 434 transpose definition TM, 433 unitary definition UM, 544 unitary is invertible theorem UMI, 546 zero definition ZM, 433 matrix addition example MA, 428 matrix components notation, 51matrix entry mathematica, 1566 sage, 1584 ti83, 1580 ti86, 1577 matrix inverse Archetype B, 518 computation theorem CINM, 516

mathematica, 1576 nonsingular matrix theorem NI, 542of a matrix inverse theorem MIMI, 521 one-sided theorem OSIS, 541 product theorem SS, 520 sage, 1590 scalar multiple theorem MISM, 523 size 2 matrices theorem TTMI, 509 transpose theorem MIT, 522 uniqueness theorem MIU, 519 matrix multiplication adjoints theorem MMAD, 483 associativity theorem MMA, 478

complex conjugation theorem MMCC, 481 definition MM, 468 distributivity theorem MMDAA, 476 entry-by-entry theorem EMP, 471identity matrix theorem MMIM, 475 inner product theorem MMIP, 480 mathematica, 1575 noncommutative example MMNC, 470 scalar matrix multiplication theorem MMSMM, 477 systems of linear equations theorem SLEMM, 462 transposes theorem MMT, 482 zero matrix theorem MMZM, 474 matrix product

as composition of linear transformations example MPMR, 1304 matrix representation basis of eigenvectors example MRBE, 1378 composition of linear transformations theorem MRCLT, 1302 definition MR, 1286 invertible theorem IMR, 1322 multiple of a linear transformation theorem MRMLT, 1301 notation, 1286 restriction to generalized eigenspace theorem MRRGE, 1514 sum of linear transformations theorem MRSLT, 1300 theorem FTMR, 1293 upper triangular theorem UTMR, 1419 matrix representations converting with change-of-basis example MRCM, 1372

example OLTTR, 1286 matrix scalar multiplication example MSM, 429 matrix vector space dimension theorem DM, 824 matrix-adjoint product eigenvalues, eigenvectors theorem EEMAP, 1915 matrix-vector product example MTV, 461 notation, 461 MBC (example), 464 MBLT (theorem), 1086 MC (notation), 51 MCC (subsection, section MO), 438 MCCN (Property), 1597 MCF (Property), 1830 MCN (definition), 1601 MCN (subsection, section CNO), 1600 MCSM (example), 564 MCT (theorem), 441 MD (chapter), 1885

ME (definition), 427ME (notation), 427ME (subsection, section PEE), 1013 ME (technique, section PT), 1626 ME (theorem), 1016ME.MMA (computation, section MMA), 1566 ME.SAGE (computation, section SAGE), 1584 ME.TI83 (computation, section TI83), 1580 ME.TI86 (computation, section TI86), 1577 MEASM (subsection, section MO), 426 MFLT (example), 1087 MI (definition), 506 MI (example), 507 MI (notation), 506 MI.MMA (computation, section MMA), 1576 MI.SAGE (computation, section SAGE), 1590 MICN (Property), 1598 MIF (Property), 1831 MIMI (theorem), 521 MINM (section), 537 MISLE (section), 503 MISM (theorem), 523 MIT (theorem), 522

MIU (theorem), 519 MIVS (example), 1274 MLT (subsection, section LT), 1083 MLTCV (theorem), 1089MLTLT (theorem), 1108 MM (definition), 468 MM (section), 460 MM (subsection, section MM), 468 MM.MMA (computation, section MMA), 1575 MR (section), 1286 MMA (section), 1566 MMA (theorem), 478 MMAD (theorem), 483 MMCC (theorem), 481 MMDAA (theorem), 476 MMEE (subsection, section MM), 471 MMIM (theorem), 475 MMIP (theorem), 480MMNC (example), 470 MMSMM (theorem), 477 MMT (theorem), 482MMZM (theorem), 474 MNEM (theorem), 1019 MNSLE (example), 463

MO (section), 426MOLT (example), 1091 more variables than equations example OSGMD, 127 theorem CMVEI, 127 MPMR (example), 1304 MR (definition), 1286 MR (notation), 1286 MRBE (example), 1378 MRCB (theorem), 1371 MRCLT (diagram), 1309 MRCLT (theorem), 1302 MRCM (example), 1372 MRLS (definition), 56 MRLS (notation), 56 MRMLT (theorem), 1301 MRRGE (theorem), 1514 MRS (subsection, section CB), 1371 MRSLT (theorem), 1300 MSCN (example), 1601 MSM (definition), 428 MSM (example), 429

MSM (notation), 428MTV (example), 461 multiplicative associativity complex numbers Property MACN, 1597 multiplicative closure complex numbers Property MCCN, 1597 field Property MCF, 1830 multiplicative commutativity complex numbers Property CMCN, 1597 multiplicative inverse complex numbers Property MICN, 1598 MVNSE (subsection, section RREF), 50 MVP (definition), 461MVP (notation), 461 MVP (subsection, section MM), 460 MVSLD (theorem), 328 MWIAA (example), 506

N (archetype), 1755 N (subsection, section O), 404 N (technique, section PT), 1620 NDMS4 (example), 1047 negation of statements technique N, 1620 NEM (theorem), 1015 NI (theorem), 542NIAO (example), 1147 NIAQ (example), 1129 NIAQR (example), 1145 NIDAU (example), 1151 nilpotent linear transformation definition NLT, 1437 NILT (diagram), 1130 NJB (theorem), 1447 NJB5 (example), 1442 NKAO (example), 1138 NLT (definition), 1437 NLT (example), 1077 NLT (section), 1435 NLT (subsection, section NLT), 1436 NLTFO (subsection, section LT), 1105 NM (definition), 173 NM (example), 174 NM (section), 172 NM (subsection, section NM), 172 NM (subsection, section OD), 1427 NM62 (example), 1439 NM64 (example), 1437 NM83 (example), 1444 NME1 (theorem), 181 NME2 (theorem), 330 NME3 (theorem), 543 NME4 (theorem), 575 NME5 (theorem), 791 NME6 (theorem), 834 NME7 (theorem), 936 NME8 (theorem), 1004 NME9 (theorem), 1328 NMI (subsection, section MINM), 537 NMLIC (theorem), 330 NMPEM (theorem), 892 NMRRI (theorem), 176 NMTNS (theorem), 179

NMUS (theorem), 180 NOILT (theorem), 1235 NOLT (definition), 1234 NOLT (notation), 1234 NOM (definition), 829 NOM (notation), 829 nonsingular columns as basis theorem CNMB, 789 nonsingular matrices linearly independent columns theorem NMLIC, 330 nonsingular matrix Archetype B example NM, 174 column space, 574elementary matrices theorem NMPEM, 892 equivalences theorem NME1, 181 theorem NME2, 330 theorem NME3, 543 theorem NME4, 575

theorem NME5, 791 theorem NME6, 834 theorem NME7, 936 theorem NME8, 1004 theorem NME9, 1328 matrix inverse, 542null space example NSNM, 179 nullity, 833 product of nonsingular matrices theorem NPNT, 538 rank theorem RNNM, 833 row-reduced theorem NMRRI, 176 trivial null space theorem NMTNS, 179 unique solutions theorem NMUS, 180 nonsingular matrix, row-reduced example NSR, 177 norm example CNSV, 404

inner product, 406 notation, 404normal matrix definition NRML, 1427 example ANM, 1428 orthonormal basis, 1434 notation A, 442 AM, 57 AME, 968 C, 1606 CCCV, **396** CCM, 438 CCN, 1598 CNA, 1596 CNE, 1596 CNM, 1596 CSM, 560 CV, 52 CVA, 201 CVC, 52 CVE, 199 CVSM, 203

D, 815	MR, 1286
DM, 894	MRLS, 56
DS, 864	MSM, 428
ELEM, 885	MVP, 461
ES, 1603	NOLT, 1234
GES, 1485	NOM, 829
GME, 969	NSM, 152
HI, 1860	NV, 404
HID, 1859	RLT, 1178
HP, 1857	RO, <mark>60</mark>
IE, 1508	ROLT, 1234
$\mathrm{IM},175$	ROM, 829
IP, 398	RREFA, 64
JB, 1441	RSM, 577
KLT, 1138	SC, 1609
LNS, 609	SE, 1604
LT, 1073	SETM, 1602
LTR, 1495	SI, 1608
M, 51	SM, 893
MA, 427	SRM, 1934
MC, 51	SSET, 1603
ME, 427	SSV, 271
MI, 506	SU, 1607

SUV, 410 T, 1845 TM. 433 VR, 1261 VSCV, 198 VSM, 426 ZCV, 53 ZM, 433 notation for a linear system example NSE, 19 NPNT (theorem), 538 NRFO (subsection, section MR), 1300 NRML (definition), 1427 NRREF (example), 65 NS.MMA (computation, section MMA), 1569 NSAO (example), 1187 NSAQ (example), 1170 NSAQR (example), 1185 NSC2A (example), 703 NSC2S (example), 703 NSC2Z (example), 702 NSDAT (example), 1193 NSDS (example), 287

NSE (example), 19 NSEAI (example), 153 NSLE (example), 56 NSLIL (example), 336 NSM (definition), 152 NSM (notation), 152 NSM (subsection, section HSE), 152 NSMS (theorem), 705 NSNM (example), 179 NSNM (subsection, section NM), 178 NSR (example), 177 NSS (example), 178 NSSLI (subsection, section LI), 331 Null space as a span example NSDS, 287 null space Archetype I example NSEAI, 153 basis theorem BNS, 334 computation example CNS1, 154

example CNS2, 156 isomorphic to kernel, 1311 linearly independent basis example LINSB, 332 mathematica, 1569 matrix definition NSM, 152 nonsingular matrix, 179 notation, 152 singular matrix, 178 spanning set example SSNS, 285 theorem SSNS, 283 subspace theorem NSMS, 705 null space span, linearly independent Archetype L example NSLIL, 336 nullity computing, 830 injective linear transformation theorem NOILT, 1235 linear transformation

definition NOLT, 1234 matrix, 829 definition NOM, 829 notation, 829, 1234 square matrix, 832 NV (definition), 404 NV (notation), 404 NVM (theorem), 1877

O (archetype), 1760 O (Property), 661 O (section), 395 OBC (subsection, section B), 792 OBNM (theorem), 1434 OBUTR (theorem), 1424 OC (Property), 206 OCN (Property), 1597 OD (section), 1414 OD (subsection, section OD), 1428 OD (theorem), 1429 OF (Property), 1831 OLTTR (example), 1286 OM (Property), 431 one

column vectors Property OC, 206 complex numbers Property OCN, 1597 field Property OF, 1831 matrices Property OM, 431 vectors Property O, 661 ONFV (example), 419 ONS (definition), 418ONTV (example), 418 orthogonal linear independence theorem OSLI, 412 set example AOS, 410set of vectors definition OSV, 409 vector pairs definition OV, 408

orthogonal vectors example TOV, 408orthonormal definition ONS, 418 matrix columns example OSMC, 548 orthonormal basis normal matrix theorem OBNM, 1434 orthonormal diagonalization theorem OD, 1429orthonormal set four vectors example ONFV, 419 three vectors example ONTV, 418 OSGMD (example), 127 OSIS (theorem), 541 OSLI (theorem), 412OSMC (example), 548 OSV (definition), 409OV (definition), 408OV (subsection, section O), 408 P (appendix), 1593 P (archetype), 1766 P (technique, section PT), 1633 particular solutions example PSHS, 258 PCNA (theorem), 1596 PCVS (example), 679 PD (section), 850 PDM (section), 919 PDM (theorem), 1936 PEE (section), 1001PEEF (theorem), 621 PI (definition), 1101 PI (subsection, section LT), 1101 PI (technique, section PT), 1627 PIP (theorem), 406PM (example), 949 PM (subsection, section EE), 949 PMI (subsection, section MISLE), 519 PMM (subsection, section MM), 474 PMR (subsection, section MR), 1310 PNLT (subsection, section NLT), 1448 POD (section), 1935

polar decomposition theorem PDM, 1936 polynomial of a matrix example PM, 949 polynomial vector space dimension theorem DP, 824 positive semi-definite creating theorem CPSM, 1879 positive semi-definite matrix definition PSM, 1879 eigenvalues theorem EPSM, 1880 practice technique P, 1633 pre-image definition PI, 1101 kernel theorem KPI, 1143 pre-images example SPIAS, 1101

principal axis theorem, 1433	CM, 430
product of triangular matrices	CMCN, 1597
theorem PTMT, 1415	CMF, 1831
Property	DCN, 1597
AA, 660	DF, 1831
AAC, 205	DMAM, 431
AACN, 1597	DSA, 660
AAF, 1831	DSAC, 206
AAM, 430	DSAM, 431
AC, 660	DVA, 660
ACC, 205	DVAC, 206
ACCN, 1596	MACN, 1597
ACF, 1830	MAF, 1831
ACM, 430	MCCN, 1597
AI, 660	MCF, 1830
AIC, 206	MICN, 1598
AICN, 1597	MIF, 1831
AIF, 1831	O, 661
AIM, 431	OC, 206
C, 660	OCN, 1597
CACN, 1597	OF, 183 1
CAF, 1830	OM, 431
CC, 205	SC, 660

SCC, 205 SCM, 430 SMA, 660 SMAC, 206 SMAM, 431 Z, 660 ZC, 206 ZCN, 1597 ZF. 1831 ZM, 431 PSHS (example), 258 PSHS (subsection, section LC), 256 PSM (definition), 1879 PSM (section), 1878 PSM (subsection, section PSM), 1879 PSM (subsection, section SD), 1032 PSMSR (theorem), 1927 PSPHS (theorem), 256 PSS (subsection, section SSLE), 20 PSSD (theorem), 857 PSSLS (theorem), 126 PT (section), 1611 PTFP (example), 1941

PTM (example), 468 PTMEE (example), 472 PTMT (theorem), 1415 Q (archetype), 1771 R (archetype), 1779 R (chapter), 1260 R.SAGE (computation, section SAGE), 1583 range full example FRAN, 1182 isomorphic to column space theorem RCSI, 1317 linear transformation example RAO, 1178 notation, 1178 of a linear transformation definition RLT, 1178 pre-image theorem RPI, 1191 subspace theorem RLTS, 1181 surjective linear transformation

theorem RSLT, 1185 via matrix representation example RVMR, 1319 rank computing theorem CRN, 830 linear transformation definition ROLT, 1233 matrix definition ROM, 829 example RNM, 829 notation, 829, 1234 of transpose example RRTI, 860 square matrix example RNSM, 832 surjective linear transformation theorem ROSLT, 1234 transpose theorem RMRT, 859 rank one decomposition size 2 example ROD2, 1891

size 4 example ROD4, 1893 theorem ROD, 1887 rank+nullity theorem RPNC, 831 RAO (example), 1178 RCLS (theorem), 120 RCSI (theorem), 1317 RD (subsection, section VS), 681 RDS (theorem), 873 READ (subsection, section B), 801 READ (subsection, section CB), 1402 READ (subsection, section CRS), 589 READ (subsection, section D), 835 READ (subsection, section DM), 905 READ (subsection, section EE), 983 READ (subsection, section FS), 644 READ (subsection, section HSE), 157 READ (subsection, section ILT), 1153 READ (subsection, section IVLT), 1244 READ (subsection, section LC), 261 READ (subsection, section LDS), 385 READ (subsection, section LI), 337

READ (subsection, section LISS), 758 READ (subsection, section LT), 1114 READ (subsection, section MINM), 552 READ (subsection, section MISLE), 524 READ (subsection, section MM), 488 READ (subsection, section MO), 444 READ (subsection, section MR), 1330 READ (subsection, section NM), 182 READ (subsection, section O), 421 READ (subsection, section PD), 874 READ (subsection, section PDM), 940 READ (subsection, section PEE), 1023 READ (subsection, section RREF), 86 READ (subsection, section S), 720READ (subsection, section SD), 1059 READ (subsection, section SLT), 1196 READ (subsection, section SS), 293 READ (subsection, section SSLE), 35 READ (subsection, section TSS), 130 READ (subsection, section VO), 208 READ (subsection, section VR), 1281 READ (subsection, section VS), 682 READ (subsection, section WILA), 13

reduced row-echelon form analysis notation, 64definition RREF, 63 example NRREF, 65 example RREF, 65 extended definition EEF, 619 notation example RREFN, 113 unique theorem RREFU, 70 reducing a span example RSC5, 366 relation of linear dependence definition RLD, 735 definition RLDCV, 316 REM (definition), 60 REMEF (theorem), 66 REMES (theorem), 61 REMRS (theorem), 580 RES (example), 382 RGEN (theorem), 1507

rings

sage, 1583 RLD (definition), 735 RLDCV (definition), 316 RLT (definition), 1178 RLT (notation), 1178 RLT (subsection, section IS), 1494 RLT (subsection, section SLT), 1178 RLTS (theorem), 1181 RMRT (theorem), 859 RNLT (subsection, section IVLT), 1233 RNM (example), 829 RNM (subsection, section D), 828 RNNM (subsection, section D), 832 RNNM (theorem), 833 RNSM (example), 832 RO (definition), 59 RO (notation), 60RO (subsection, section RREF), 59 ROD (section), 1887ROD (theorem), 1887 ROD2 (example), 1891 ROD4 (example), 1893

ROLT (definition), 1233 ROLT (notation), 1234 ROM (definition), 829 ROM (notation), 829 ROSLT (theorem), 1234 row operations definition RO, 59 elementary matrices, 885, 886 notation, 60row reduce mathematica, 1567 sage, 1586 ti83, 1581 ti86, 1578 row space Archetype I example RSAI, 577 as column space, 586basis example RSB, 785 theorem BRS, 583 matrix, 577 notation, 577

row-equivalent matrices theorem REMRS, 580 subspace theorem RSMS, 719 row-equivalent matrices definition REM, 60 example TREM, 60 row space, 580 row spaces example RSREM, 582 theorem REMES, 61 row-reduce the verb definition RR, 85 row-reduced matrices theorem REMEF, 66 RPI (theorem), 1191 RPNC (theorem), 831 RPNDD (theorem), 1235 RR (definition), 85 RR.MMA (computation, section MMA), 1567 RR.SAGE (computation, section SAGE), 1586 RT (subsection, section PD), 859 RR.TI83 (computation, section TI83), 1581

RR.TI86 (computation, section TI86), 1578 RREF (definition), 63 RREF (example), 65 RREF (section), 50RREF (subsection, section RREF), 63 RREFA (notation), 64 RREFN (example), 113 RREFU (theorem), 70 RRTI (example), 860 RS (example), 787 RSAI (example), 577 RSB (example), 785 RSC4 (example), 380 RSC5 (example), 366 RSLT (theorem), 1185 RSM (definition), 577 RSM (notation), 577 RSM (subsection, section CRS), 576 RSMS (theorem), 719 RSNS (example), 706 RSREM (example), 582 RVMR (example), 1319

```
S (archetype), 1785
S (definition), 694
S (example), 174
S (section), 694
SAA (example), 81
SAB (example), 79
SABMI (example), 503
SAE (example), 83
sage
   eigenspaces (computation), 1591
   linear solve (computation), 1588
   matrix entry (computation), 1584
   matrix inverse (computation), 1590
   rings (computation), 1583
   row reduce (computation), 1586
   transpose of a matrix (computation), 1591
   vector linear combinations (computation),
       1589
SAGE (section), 1583
SAN (example), 1188
SAR (example), 1172
SAS (section), 1949
SAV (example), 1175
```

```
SC (definition), 1608
SC (example), 1609
SC (notation), 1609
SC (Property), 660
SC (subsection, section S), 719
SC (subsection, section SET), 1606
SC3 (example), 694
SCAA (example), 275
SCAB (example), 279
SCAD (example), 290
scalar closure
   column vectors
     Property SCC, 205
   matrices
     Property SCM, 430
   vectors
     Property SC, 660
scalar multiple
   matrix inverse, 523
scalar multiplication
   zero scalar
     theorem ZSSM, 676
   zero vector
```

theorem ZVSM, 677 zero vector result theorem SMEZV, 679 scalar multiplication associativity column vectors Property SMAC, 206 matrices Property SMAM, 431 vectors Property SMA, 660 SCB (theorem), 1377 SCC (Property), 205 SCM (Property), 430 SD (section), 1029SDS (example), 865 SE (definition), 1604SE (notation), 1604secret sharing 6 ways example SS6W, 1952 SEE (example), 946 SEEF (example), 619 SER (theorem), 1033

set

cardinality definition C, 1606 example CS, 1606 notation, 1606 complement definition SC, 1608 example SC, 1609 notation, 1609 definition SET, 1602 empty definition ES, 1603 equality definition SE, 1604notation, 1604 intersection definition SI, 1608 example SI, 1608 notation, 1608membership example SETM, 1602 notation, 1602size, 1606

subset, 1603 union definition SU, 1607 example SU, 1607 notation, 1607SET (definition), 1602SET (section), 1602SETM (example), 1602 SETM (notation), 1602 shoes, 520SHS (subsection, section HSE), 146 SI (definition), 1608 SI (example), 1608SI (notation), 1608SI (subsection, section IVLT), 1229 SIM (definition), 1029 similar matrices equal eigenvalues example EENS, 1036 eual eigenvalues theorem SMEE, 1035 example SMS3, 1031 example SMS5, 1030

similarity definition SIM, 1029 equivalence relation theorem SER, 1033 singular matrix Archetype A example S, 174 null space example NSS, 178 singular matrix, row-reduced example SRR, 176 singular value decomposition theorem SVD, 1924 singular values definition SV, 1923 SLE (chapter), 2 SLE (definition), 18 SLE (subsection, section SSLE), 16 SLELT (subsection, section IVLT), 1240 SLEMM (theorem), 462 SLSLC (theorem), 229 SLT (definition), 1169 SLT (section), 1169

- SLTB (theorem), 1192 SLTD (subsection, section SLT), 1193 SLTD (theorem), 1193 SLTLT (theorem), 1106 SM (definition), 893 SM (notation), 893 SM (subsection, section SD), 1029 SM2Z7 (example), 1836 SM32 (example), 714 SMA (Property), 660 SMAC (Property), 206 SMAM (Property), 431 SMEE (theorem), 1035 SMEZV (theorem), 679 SMLT (example), 1109 SMS (theorem), 435SMS3 (example), 1031 SMS5 (example), 1030SMZD (theorem), 935 SMZE (theorem), 1003SNCM (theorem), 543 SO (subsection, section SET), 1607 socks, 520
- SOL (subsection, section B), 805SOL (subsection, section CB), 1406 SOL (subsection, section CRS), 601 SOL (subsection, section D), 842SOL (subsection, section DM), 912 SOL (subsection, section EE), 989 SOL (subsection, section F), 1844 SOL (subsection, section FS), 650SOL (subsection, section HSE), 164 SOL (subsection, section ILT), 1160 SOL (subsection, section IVLT), 1251 SOL (subsection, section LC), 267 SOL (subsection, section LDS), 389 SOL (subsection, section LI), 348 SOL (subsection, section LISS), 764 SOL (subsection, section LT), 1120 SOL (subsection, section MINM), 556 SOL (subsection, section MISLE), 531 SOL (subsection, section MM), 496 SOL (subsection, section MO), 453 SOL (subsection, section MR), 1337 SOL (subsection, section NM), 188 SOL (subsection, section O), 424

SOL (subsection, section PD), 879 SOL (subsection, section PDM), 943 SOL (subsection, section PEE), 1026 SOL (subsection, section RREF), 97 SOL (subsection, section S), 726SOL (subsection, section SD), 1063 SOL (subsection, section SLT), 1203 SOL (subsection, section SS), 301SOL (subsection, section SSLE), 43SOL (subsection, section T), 1855 SOL (subsection, section TSS), 138 SOL (subsection, section VO), 215 SOL (subsection, section VR), 1284 SOL (subsection, section VS), 688 SOL (subsection, section WILA), 15 solution set Archetype A example SAA, 81 archetype E example SAE, 83 theorem PSPHS, 256 solution set of a linear system definition SSSLE, 19

solution sets possibilities theorem PSSLS, 126 solution to a linear system definition SSLE, 18 solution vector definition SOLV, 55 SOLV (definition), 55 solving homogeneous system Archetype A example HISAA, 149 Archetype B example HUSAB, 148 Archetype D example HISAD, 150 solving nonlinear equations example STNE, 16 SP4 (example), 700 span basic example ABS, 271 basis theorem BS, 376

definition SS, 709 definition SSCV, 270 improved example IAS, 585 notation, 271 reducing example RSC4, 380 reduction example RS, 787 removing vectors example COV, 370 reworking elements example RES, 382 set of polynomials example SSP, 712 subspace theorem SSS, 709 span of columns Archetype A example SCAA, 275 Archetype B example SCAB, 279 Archetype D

example SCAD, 290 spanning set crazy vector space example SSC, 751 definition TSVS, 746 matrices example SSM22, 749 more vectors theorem SSLD, 816 polynomials example SSP4, 746 SPIAS (example), 1101 SQM (definition), 173 square root eigenvalues, eigenspaces theorem EESR, 1930 matrix definition SRM, 1934 notation, 1934 positive semi-definite matrix theorem PSMSR, 1927 unique theorem USR, 1933

SR (section), 1926SRM (definition), 1934 SRM (notation), 1934 SRM (subsection, section SR), 1927 SRR (example), 176 SS (definition), 709 SS (example), 893 SS (section), 270SS (subsection, section LISS), 745 SS (theorem), 520 SS6W (example), 1952SSC (example), 751 SSCV (definition), 270 SSET (definition), 1603 SSET (example), 1603 SSET (notation), 1603 SSLD (theorem), 816 SSLE (definition), 18 SSLE (section), 16 SSM22 (example), 749 SSNS (example), 285 SSNS (subsection, section SS), 283 SSNS (theorem), 283

SSP (example), 712 SSP4 (example), 746 SSRLT (theorem), 1188 SSS (theorem), 709 SSSLE (definition), 19 SSSLT (subsection, section SLT), 1188 SSV (notation), 271 SSV (subsection, section SS), 270 standard unit vector notation, 410 starting proofs technique GS, 1617 STLT (example), 1107 STNE (example), 16 SU (definition), 1607 SU (example), 1607 SU (notation), 1607submatrix notation, 893 subset definition SSET, 1603 notation, 1603 subspace

as null space example RSNS, 706 characterized example ASC, 1273 definition S, 694 in P_4 example SP4, 700 not, additive closure example NSC2A, 703 not, scalar closure example NSC2S, 703 not, zero vector example NSC2Z, 702 testing theorem TSS, 698 trivial definition TS, 704 verification example SC3, 694 example SM32, 714 subspaces equal dimension theorem EDYES, 858 surjective Archetype N example SAN, 1188 example SAR, 1172 not example NSAQ, 1170 example NSAQR, 1185 not, Archetype O example NSAO, 1187 not, by dimension example NSDAT, 1193 polynomials to matrices example SAV, 1175 surjective linear transformation bases theorem SLTB, 1192 surjective linear transformations dimension theorem SLTD, 1193 SUV (definition), 409SUV (notation), 410 SUVB (theorem), 779 SUVOS (example), 410

SV (definition), 1923SVD (section), 1914 SVD (subsection, section SVD), 1923 SVD (theorem), 1924 SVP4 (example), 856 SYM (definition), 434 SYM (example), 435 symmetric matrices theorem SMS, 435 symmetric matrix example SYM, 435 system of equations vector equality example VESE, 200 system of linear equations definition SLE, 18

T (archetype), 1790 T (definition), 1845 T (notation), 1845 T (part), 1829 T (section), 1845 T (technique, section PT), 1613 TCSD (example), 902 TD (section), 1897 TD (subsection, section TD), 1897 TD (theorem), 1898 TD4 (example), 1902 TDEE (theorem), 1908 TDEE6 (example), 1911 TDSSE (example), 1905 TDSSE (subsection, section TD), 1904 technique

C, 1618 CD, 1623 CP, 1621 CV, 1622 D, 1611 DC, 1628 E, 1619 GS, 1617 I, 1630 L, 1614 LC, 1634 ME, 1626 N, 1620

	P, 1633	CHT, 1563
	PI, 1627	CILTI, 1152
	T, 1613	CINM, 516
	U, 1625	CIVLT, 1227
theorem		CLI, 1275
	AA, 443	CLTLT, 1111
	AIP, 485	CMVEI, 127
	AISM, 678	CNMB, 789
	AIU, 675	COB, 793
	AMA, 442	CPSM, 1879
	AMSM, 443	CRMA, 439
	BCS, 569	CRMSM, 440
	BIS, 823	CRN, 830
	BNS, 334	CRSM, 397
	BRS, 583	CRVA, 396
	BS, 376	$\mathrm{CSCS}, 563$
	CB, 1359	CSLTS, 1195
	CCM, 440	CSMS, 719
	CCRA, 1599	CSNM, 574
	CCRM, 1599	CSRN, 123
	CCT, 1600	CSRST, 586
	CFDVS, 1272	CSS, 1276
	CFNLT, 1458	CUMOS, 547

DC, 1039 DCM, 824 DCP, 1013 DEC, 901 DED, 1048 DEM, 932 DEMMM, 933 DER, 897 DERC, 923 DFS, 861 DGES, 1531 DIM, 930 DLDS, 365 DM, 824 DMFE, 1044 DMHP, 1863 DMMP, 1865 DMST, 896 DNLT, 1449 DP, 824 DRCM, 922 DRCMA, 924 DRCS, 920

DRMM, 938 DSD, 872 DSFB, 865 DSFOS, 867 DSLI, 870 DSZI, 869 DSZV, 868 DT, 899 DVM, 1871 DZRC, 919 EDELI, 1001 EDYES, 858 EEMAP, 1915 EER, 1384 EESR, 1930 EIM, 1009 EIS, 1479 ELIS, 850 EMDRO, 886 EMHE, 953 EMMVP, 466 EMN, 891 EMNS, 965

EMP, **471** EMRCP, 962 EMS. 963 ENLT, 1448 EOMP, 1006 EOPSS, 24 EPM, 1007 EPSM, 1880 ERMCP, 1012 ESMM, 1005 ETM, 1011 FIMP, 1833 FS, 625 FTMR, 1293 FVCS, 124 G, 852 GEK, 1487 GESD, 1518 GESIS, 1485 GSP, **413** HMIP, 486 HMOE, 1022 HMRE, 1020

HMVEI, 151 HPC, 1858 HPDAA, 1861 HPHI, 1860 HPHID, 1859 HPSMM, 1862 HSC, 147 ICBM, 1360 ICLT, 1227 IFDVS, 1274 IILT, 1220 ILTB, 1149 ILTD, 1150 ILTIS, 1221 ILTLI, 1148 ILTLT, 1219 IMILT, 1327 IMR, 1322 IP, 1940 IPAC, 402 IPN, 406 IPSM, **401** IPVA, 400

ISRN, 122 ITMT, 1417 IVSED, 1232 **JCFLT**, 1533 KILT, 1144 KLTS, 1140 KNSI, 1311 KPI, 1143 KPIS, 1481 KPLT, 1450 KPNLT, 1453 LIVHS, 321 LIVRN, 325 LNSMS, 720 LSMR, 1944 LTDB, 1094 LTLC, 1093 LTTZZ, 1080 MBLT, 1086 MCT, 441 ME, 1016 MIMI, 521 MISM, 523

MIT, 522 MIU, 519 MLTCV, 1089 MLTLT, 1108 MMA, 478 MMAD, 483 MMCC, 481 MMDAA, 476 MMIM, 475 MMIP, 480 MMSMM, 477 MMT, 482 MMZM, 474 MNEM, 1019 MRCB, 1371 MRCLT, 1302 MRMLT, 1301 MRRGE, 1514 MRSLT, 1300 MVSLD, 328 NEM, 1015 NI, 542 NJB, 1447

NME1, 181 NME2, 330 NME3, 543 NME4, 575 NME5, 791 NME6, 834 NME7, 936 NME8, 1004 NME9, 1328 NMLIC, **330** NMPEM, 892 NMRRI, 176 NMTNS, 179 NMUS, 180 NOILT, 1235 NPNT, 538 NSMS, 705 NVM, 1877 OBNM, 1434 OBUTR, 1424 OD, 1429 OSIS, 541 OSLI, 412

PCNA, 1596 PDM, 1936 PEEF, 621 PIP, 406 PSMSR, 1927 PSPHS, 256 PSSD, 857 **PSSLS**, 126 PTMT, 1415 RCLS, 120 RCSI, 1317 RDS. 873 REMEF, 66 REMES, 61 REMRS, 580 RGEN, 1507 RLTS, 1181 RMRT, 859 RNNM, 833 ROD, 1887 ROSLT, 1234 RPI, 1191 RPNC, 831

RPNDD, 1235 RREFU, 70 RSLT, 1185 RSMS, 719 SCB, 1377 SER, 1033 SLEMM, 462 SLSLC, 229 SLTB, 1192 SLTD, 1193 SLTLT, 1106 SMEE, 1035 SMEZV, 679 SMS, 435 SMZD, 935 SMZE, 1003 SNCM, 543 SS, 520 SSLD, 816 SSNS, 283 SSRLT, 1188 SSS, 709 SUVB, 779

SVD, 1924 TD, 1898 TDEE, 1908 technique T, 1613 TIST, 1848 TL, 1846 TMA, 436 TMSM, 436 TSE, 1849 TSRM, 1847 TSS, 698 TT, 437 TTMI, 509 UMCOB, 798 UMI, 546 UMPIP, **549** USR, 1933 UTMR, 1419 VFSLS, 244 VRI, 1270 VRILT, 1272 VRLT, 1262 VRRB, 755

VRS, 1271 VSLT, 1110 VSPCV, 205 VSPM, 430 ZSSM, 676 ZVSM, 677 ZVU, 675

ti83

matrix entry (computation), 1580 row reduce (computation), 1581 vector linear combinations (computation), 1582 TI83 (section), 1580 ti86 matrix entry (computation), 1577 row reduce (computation), 1578 transpose of a matrix (computation), 1580 vector linear combinations (computation), 1579 TI86 (section), 1577 TIS (example), 1476 TIST (theorem), 1848 TIVS (example), 1273

TKAP (example), 1141 TL (theorem), 1846TLC (example), 222 TM (definition), 433 TM (example), 433 TM (notation), 433 TM (subsection, section OD), 1415 TM.MMA (computation, section MMA), 1575 TM.SAGE (computation, section SAGE), 1591 TM.TI86 (computation, section TI86), 1580 TMA (theorem), 436 TMP (example), 6TMSM (theorem), 436 TOV (example), 408trace definition T, 1845 linearity theorem TL, 1846 matrix multiplication theorem TSRM, 1847 notation, 1845 similarity theorem TIST, 1848

sum of eigenvalues theorem TSE, 1849 trail mix example TMP, 6 transpose matrix scalar multiplication theorem TMSM, 436 example TM, 433 matrix addition theorem TMA, 436 matrix inverse, 521, 522 notation, 433 scalar multiplication, 437 transpose of a matrix mathematica, 1575 sage, 1591 ti86, 1580 transpose of a transpose theorem TT, 437 TREM (example), 60triangular decomposition entry by entry, size 6 example TDEE6, 1911

entry by entry theorem TDEE, 1908 size 4 example TD4, 1902 solving systems of equations example TDSSE, 1905 theorem TD, 1898 triangular matrix inverse theorem ITMT, 1417 trivial solution system of equations definition TSHSE, 148 TS (definition), 704 TS (subsection, section S), 698 TSE (theorem), 1849TSHSE (definition), 148 TSM (subsection, section MO), 433 TSRM (theorem), 1847 TSS (section), 112TSS (subsection, section S), 707 TSS (theorem), 698TSVS (definition), 746

TT (theorem), 437 TTMI (theorem), 509 TTS (example), 21 typical systems, 2×2 example TTS, 21

U (archetype), 1794 U (technique, section PT), 1625 UM (definition), 544 UM (subsection, section MINM), 544 UM3 (example), 545 UMCOB (theorem), 798 UMI (theorem), 546 UMPIP (theorem), 549 unique solution, 3×3 example US, 29example USR, 62 uniqueness technique U, 1625 unit vectors basis theorem SUVB, 779 definition SUV, 409

orthogonal example SUVOS, 410 unitary permutation matrix example UPM, 545 size 3 example UM3, 545 unitary matrices columns theorem CUMOS, 547 unitary matrix inner product theorem UMPIP, 549 UPM (example), 545 upper triangular matrix definition UTM, 1415 US (example), 29 USR (example), 62 USR (theorem), 1933 UTM (definition), 1415 UTMR (subsection, section OD), 1418 UTMR (theorem), 1419

V (archetype), 1799 V (chapter), 197 VA (example), 202 Vandermonde matrix definition VM, 1870 vandermonde matrix determinant theorem DVM, 1871nonsingular theorem NVM, 1877 size 4 example VM4, 1870 VEASM (subsection, section VO), 199 vector addition definition CVA, 201 column definition CV, 52 equality definition CVE, 199 notation, 199 inner product definition IP, 398

norm definition NV, 404notation, 52of constants definition VOC, 54 product with matrix, 461, 468 scalar multiplication definition CVSM, 203 vector addition example VA, 202vector component notation, 52 vector form of solutions Archetype D example VFSAD, 232 Archetype I example VFSAI, 248 Archetype L example VFSAL, 252 example VFS, 237 mathematica, 1571 theorem VFSLS, 244 vector linear combinations

mathematica, 1568 sage, 1589 ti83, 1582 ti86, 1579 vector representation example AVR, 754 example VRC4, 1263 injective theorem VRI, 1270 invertible theorem VRILT, 1272 linear transformation definition VR, 1261 notation, 1261 theorem VRLT, 1262 surjective theorem VRS, 1271 theorem VRRB, 755 vector representations polynomials example VRP2, 1267 vector scalar multiplication example CVSM, 204

vector space characterization theorem CFDVS, 1272 column vectors definition VSCV, 198 definition VS, 659 infinite dimension example VSPUD, 828 linear transformations theorem VSLT, 1110 over integers mod 5 example VSIM5, 1834 vector space of column vectors notation, 198 vector space of functions example VSF, 667 vector space of infinite sequences example VSIS, 667 vector space of matrices definition VSM, 426 example VSM, 663 notation, 426vector space of polynomials

example VSP, 665 vector space properties column vectors theorem VSPCV, 205 matrices theorem VSPM, 430 vector space, crazy example CVS, 670 vector space, singleton example VSS, 668 vector spaces isomorphic definition IVS, 1230 theorem IFDVS, 1274 VESE (example), 200 VFS (example), 237 VFSAD (example), 232 VFSAI (example), 248 VFSAL (example), 252 VFSLS (theorem), 244 VFSS (subsection, section LC), 232 VFSS.MMA (computation, section MMA), 157 IVS (section), 659 VLC.MMA (computation, section MMA), 1568 VS (subsection, section VS), 659

VLC.SAGE (computation, section SAGE), 1589 VLC.TI83 (computation, section TI83), 1582 VLC.TI86 (computation, section TI86), 1579 VM (definition), 1870 VM (section), 1870VM4 (example), 1870 VO (section), 198VOC (definition), 54 VR (definition), 1261 VR (notation), 1261VR (section), 1261VR (subsection, section LISS), 754 VRC4 (example), 1263 VRI (theorem), 1270 VRILT (theorem), 1272 VRLT (theorem), 1262 VRP2 (example), 1267 VRRB (theorem), 755 VRS (theorem), 1271 VS (chapter), 658VS (definition), 659

VSCV (definition), 198 VSCV (example), 663 VSCV (notation), 198 VSF (example), 667 VSIM5 (example), 1834 VSIS (example), 667 VSLT (theorem), 1110 VSM (definition), 426 VSM (example), 663 VSM (notation), 426VSP (example), 665 VSP (subsection, section MO), 430 VSP (subsection, section VO), 205 VSP (subsection, section VS), 674 VSPCV (theorem), 205 VSPM (theorem), 430 VSPUD (example), 828 VSS (example), 668 W (archetype), 1803WILA (section), 3

X (archetype), 1808

Z (Property), 660 ZC (Property), 206 ZCN (Property), 1597 ZCV (definition), 53 ZCV (notation), 53 zero complex numbers Property ZCN, 1597 field Property ZF, 1831 zero column vector definition ZCV, 53 notation, 53 zero matrix notation, 433 zero vector column vectors Property ZC, 206 matrices Property ZM, 431 unique theorem ZVU, 675 vectors

Property Z, 660 ZF (Property), 1831 ZM (definition), 433 ZM (notation), 433 ZM (Property), 431 ZNDAB (example), 936 ZSSM (theorem), 676 ZVSM (theorem), 677 ZVU (theorem), 675