A First Course in Linear Algebra

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

Version 0.30
January 18, 2005

Preface

This textbook is designed to teach the university mathematics student the basics of the subject of linear algebra. 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 of developing the definitions and theorems of 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 under the terms of the GNU Free Documentation License (GFDL), which gives you more rights than most copyrights. Loosely speaking, you may make as many copies as you like at no cost, and you may distribute these 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, you must also license the new version with the GFDL. So the book has lots of inherent freedom, and no one is allowed to distribute a modified version that restricts these freedoms. (See the license itself for all the exact details of the rights you have been granted.)

Notice that initially most people are struck by the notion that this book is **free** (the French would say *gratis*, 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 updates be designed 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 all to benefit.

Topics The first half of this text (through Chapter M [159]) is basically a course in matrix algebra, though the foundation of some more advanced ideas is also being laid in these early sections. Vectors are presented exclusively as column vectors (since we also have the typographic freedom to avoid the cost-cutting move of displaying column vectors inline as the transpose of row vectors), and linear combinations are presented very early. Spans, null spaces and ranges are also presented very 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 late.

However, with a definition built on linear combinations of column vectors, it should seem more natural than the usual 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 vectors and matrices are first defined, but the notion of a vector space is saved for a more axiomatic treatment later. Once bases and dimension have been explored in the context of vector spaces, linear transformations and their matrix representations follow. The goal of the book is to go as far as canonical forms and matrix decompositions in the Core, with less central topics collected in a section of Topics.

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, 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 advice mostly.

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

- Math 232 is the post-calculus linear algebra course taught at the University of Puget Sound 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, Chapter A [469].) 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 book.

- 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. 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 TEX 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 TEX 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.

However, much of my motivation for writing this book is captured 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 Chapter, 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, but there is also a Section B. 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, 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. 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 the most general representations of linear transformations (Jordan and rational canonical forms) and perhaps matrix decompositions (LU, QR, singular value). Of course, there will not be universal agreement on what should, or should not, constitute the Core, but the main idea will be to limit it to about forty sections. Topics 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 should illustrate the power and widespread applicability of linear algebra to as many fields as possible. The Archetypes (Chapter A [469]) 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, they are a repository that contains enough variety to provide abundant examples of key theorems, while also providing counterexamples to hypotheses or converses of theorems.

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.

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, over time solutions will also be provided. 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.T34 is a solution to an exercise in Section MO that is theoretical in nature. The number '34' has no intrinsic meaning.

More on Freedom This book is freely-distributable under the terms of the GFDL, along with the underlying TEX 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. Students will not feel the need to sell back their book, and in future years can even pick up a newer edition freely.

- 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.

- 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 TEX, 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 TEX 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 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. Does there need to 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 gains, 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 chapters 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 `linear.ups.edu`).

Robert A. Beezer
Tacoma, Washington
December, 2004

Contents

Definitions

Theorems

Notation

Section CNO

Examples

Proof Techniques

Computation Notes

Contributors

Beezer, David. St. Charles Borromeo School. \relax

Beezer, Robert. University of Puget Sound. http://buzzard.ups.edu/

Riegsecker, Joe. Middlebury, Indiana. joepye(at)pobox(dot)com

Phelps, Douglas. University of Puget Sound. \relax

GNU Free Documentation License

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 non-commercially. 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 C
Core

SLE: Systems of Linear Equations

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:

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

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 [547]). 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 A
An application: packaging trail mix

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 much 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 a 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 (kg/batch)	Peanuts (kg/batch)	Chocolate (kg/batch)	Cost ($/kg)	Sale Price ($/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

$$b \geq 0 \qquad\qquad s \geq 0 \qquad\qquad f \geq 0.$$

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

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

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

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

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,

$$300(4.99 - 3.69) + 300(5.50 - 3.86) + 900(6.50 - 4.45) = 2727$$

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 (kg/batch)	Peanuts (kg/batch)	Chocolate (kg/batch)	Cost (\$/kg)	Sale Price (\$/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

$$b \geq 0, \qquad s \geq 0, \qquad f \geq 0$$

and

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

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. Grab your pencil and paper and play along.

$$\begin{aligned}
\frac{7}{15}(4f-3300) + \frac{6}{15}(-5f+4800) + \frac{2}{15}f &= 0f + \frac{5700}{15} = 380\\
\frac{5}{15}(4f-3300) + \frac{5}{15}(-5f+4800) + \frac{5}{15}f &= 0f + \frac{7500}{15} = 500\\
\frac{3}{15}(4f-3300) + \frac{4}{15}(-5f+4800) + \frac{8}{15}f &= 0f + \frac{9300}{15} = 620
\end{aligned}$$

Again, right now, do not be concerned about how you might derive expressions like those for b and s that fit so nicely into this system of equations. But do convince yourself that they lead to 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

$$b \geq 0 \qquad \Rightarrow \qquad 4f-3300 \geq 0 \qquad \Rightarrow \qquad f \geq 825.$$

Similarly for the standard mix,

$$s \geq 0 \qquad \Rightarrow \qquad -5f+4800 \geq 0 \qquad \Rightarrow \qquad f \leq 960.$$

So, as production manager, you really have to choose a value of f from the 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 will *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 lead 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 $0.21(960) + 2463 = 2664.6$. 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, system science or management science. More specifically, this is an 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 simultaneous 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. Explain the importance of the procedures described in the trail mix application (Subsection WILA.A [4]) from the point-of-view of the production manager.

Subsection EXC
Exercises

M70 Contributed by Robert Beezer
In Example TMP [4] 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. Solution [13]

Subsection SOL
Solutions

M70 Exercise [11] Contributed by Robert Beezer

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 $\{825, 826, \ldots, 960\}$. 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 SSSLE
Solving Systems of Simultaneous 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.

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

$$\begin{aligned} x^2 + y^2 &= 1 \\ -x + \sqrt{3}y &= 0 \end{aligned}$$

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 the only ones. We like to write solutions as sets, so in this case we write the set of solutions as

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

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. To help you in this process, we will digress, at irregular intervals, about some important aspect of working with 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). 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, 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"? ♢

Definition SSLE
System of Simultaneous Linear Equations
A **system of simultaneous linear equations** is a collection of m equations in the variable quantities x_1, x_2, $x_3, \ldots, x_n$ of the form,

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

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 [547], but these facts will not be critical until we reach Section O [145]. For now, here is an example to illustrate using the notation introduced in Definition SSLE [16].

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

$$\begin{aligned} x_1 + 2x_2 + \quad x_4 &= 7 \\ x_1 + x_2 + x_3 - x_4 &= 3 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

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

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

Additionally, convince yourself that $x_1 = -2$, $x_2 = 4$, $x_3 = 2$, $x_4 = 1$ is one solution (but it is not the only one!). ⊚

We will often shorten the term "system of simultaneous linear equations" to "system of linear equations" or just "system of equations" leaving the linear or simultaneous aspects implied.

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,

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ x_1 - x_2 &= 4 \end{aligned}$$

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,

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ 4x_1 + 6x_2 &= 6. \end{aligned}$$

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 [42], Theorem VFSLS [104]). Notice now how the second equation is just a multiple of the first.

One more minor adjustment

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ 4x_1 + 6x_2 &= 10. \end{aligned}$$

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 simultaneous linear equations has a solution set that is empty, contains a single solution or contains infinitely many solutions (Theorem PSSLS [58]). Example STNE [15] yielded exactly two solutions, but this does not contradict the forthcoming theorem, since the equations are not linear and do not match the form of Definition SSLE [16].

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 ES
Equivalent Systems
Two systems of simultaneous linear equations are **equivalent** if their solution sets are equal. △

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

1. Swap the locations of two equations in the list.
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. △

These descriptions might seem a bit vague, but the proof or the examples that follow should make it clear what is meant by each.

Proof Technique T
Theorems
Higher mathematics is about understanding theorems. Reading them, understanding them, applying them, proving them. We are ready to prove our first momentarily. 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 something-else-happens." The "something-happens" part is the **hypothesis** and the "something-else-happens" 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. ◊

Theorem EOPSS
Equation Operations Preserve Solution Sets
Suppose we apply one of the three equation operations of Definition EO [19] to the system

of simultaneous linear equations

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

Then the original system and the transformed system are equivalent systems. □

Proof Before we begin the proof, we make two comments about proof techniques.

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 [19], rewrite the statement of the theorem in an "if-then" 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 SE
Set Equality
In the theorem we are trying to prove, the conclusion is that two systems are equivalent. By Definition ES [18] 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. So lets add it to our toolbox now.

A **set** is just a collection of items, which we refer to generically as **elements**. If A is a set, and a is one of its elements, we write that piece of information as $a \in A$. Similarly, if b is not in A, we write $b \notin A$. Given two sets, A and B, we say that A is a **subset** of B if all the elements of A are also in B. More formally (and much easier to work with) we describe this situation as follows: A is a subset of B if whenever $x \in A$, then $x \in B$. Notice the use of the "if-then" construction here. The notation for this is $A \subseteq B$. (If we want to disallow the possibility that A is the same as B, we use $A \subset B$.)

But 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. ◊

Now we can take each equation operation in turn and show that the solution sets of the two systems are equal, using the technique just outlined.

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,

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1\\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2\\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3\\
&\vdots\\
\alpha a_{i1}x_1 + \alpha a_{i2}x_2 + \alpha a_{i3}x_3 + \cdots + \alpha a_{in}x_n &= \alpha b_i\\
&\vdots\\
a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m.
\end{aligned}$$

   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 every other equation of the transformed system true. We also know that

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

which we can multiply by α to get

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

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 every other equation of the original system true. We also know that

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

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

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

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,

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1\\
&a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2\\
&a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n = b_3\\
&\vdots\\
&(\alpha a_{i1} + a_{j1})x_1 + (\alpha a_{i2} + a_{j2})x_2 + \cdots + (\alpha a_{in} + a_{jn})x_n = \alpha b_i + b_j\\
&\vdots\\
&a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m.
\end{aligned}$$

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 it makes every other equation of the transformed system true. Using the fact that it makes the i-th and j-th equations true, we find

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

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 every other equation of the original system true. We then find

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

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. ■

Theorem EOPSS [20] 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.

$$
\begin{aligned}
x_1 + 2x_2 + 2x_3 &= 4 \\
x_1 + 3x_2 + 3x_3 &= 5 \\
2x_1 + 6x_2 + 5x_3 &= 6
\end{aligned}
$$

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

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 2x_1 + 6x_2 + 5x_3 &= 6 \end{aligned}$$

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

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 0x_1 + 2x_2 + 1x_3 &= -2 \end{aligned}$$

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

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 0x_1 + 0x_2 - 1x_3 &= -4 \end{aligned}$$

$\alpha = -1$ times equation 3:

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 0x_1 + 0x_2 + 1x_3 &= 4 \end{aligned}$$

which can be written more clearly as

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ x_2 + x_3 &= 1 \\ x_3 &= 4 \end{aligned}$$

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 [20]. 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 systems of equations). ⊚

Example IS
Three equations, infinitely many solutions
The following system of equations made an appearance earlier in this section (Example NSE [17]), where we listed *one* of its solutions. Now, we will try to find all of the

solutions to this system.

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ x_1 + x_2 + x_3 - x_4 &= 3 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

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

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 - x_2 + x_3 - 2x_4 &= -4 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

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

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 - x_2 + x_3 - 2x_4 &= -4 \\ 0x_1 - 5x_2 + 5x_3 - 10x_4 &= -20 \end{aligned}$$

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

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 - x_2 + x_3 - 2x_4 &= -4 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 &= 0 \end{aligned}$$

$\alpha = -1$ times equation 2:

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 + x_2 - x_3 + 2x_4 &= 4 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 &= 0 \end{aligned}$$

which can be written more clearly as

$$\begin{aligned} x_1 + 2x_2 + x_4 &= 7 \\ x_2 - x_3 + 2x_4 &= 4 \\ 0 &= 0 \end{aligned}$$

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 just need to only 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. Lets choose x_3 and x_4 to be *anything* we please, say $x_3 = \beta_3$ and $x_4 = \beta_4$. Then equation 2 becomes

$$x_2 - \beta_3 + 2\beta_4 = 4 \qquad \Rightarrow \qquad x_2 = 4 + \beta_3 - 2\beta_4$$

Now we can take these arbitrary values for x_3 and x_4, and this expression for x_2 and employ them in equation 1,

$$x_1 + 2(4 + \beta_3 - 2\beta_4) + \beta_4 = 7 \qquad \Rightarrow \qquad x_1 = -1 - 2\beta_3 + 3\beta_4$$

So our arbitrary choices of values for x_3 and x_4 (β_3 and β_4) translate into specific values of x_1 and x_2. The lone solution given in Example NSE [17] was obtained by choosing $\beta_3 = 2$ and $\beta_4 = 1$. Now we can easily and quickly find many more (infinitely more). Suppose we choose $\beta_3 = 5$ and $\beta_4 = -2$, 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

$$S = \{\,(-1 - 2\beta_3 + 3\beta_4,\ 4 + \beta_3 - 2\beta_4,\ \beta_3,\ \beta_4) \mid \beta_3 \in \mathbb{C},\ \beta_4 \in \mathbb{C}\}$$

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. ⊚

In the next section we will describe how to use equation operations to systematically solve any system of simultaneous linear equations. But first, one of our more important pieces of advice about doing mathematics.

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 hard to speak mathematics, and so that is the topic of this technique.

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 [15]). 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, 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 perhaps this advice will be even clearer, since you know that a compiler will often complain with an error message if you confuse your objects.

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 slooooowly 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.

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! $\diamond$

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 Contributed by Robert Beezer
Find a solution to the system in Example IS [24] where $\beta_3 = 6$ and $\beta_4 = 2$. Find two other solutions to the system. Find a solution where $\beta_1 = -17$ and $\beta_2 = 14$. How many possible answers are there to each of these questions?

C20 Contributed by Robert Beezer
Each archetype (Chapter A [**??**]) 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 [473], Archetype B [478], Archetype C [483], Archetype D [487], Archetype E [491], Archetype F [495], Archetype G [501], Archetype H [505], Archetype I [510], and Archetype J [515].

M30 Contributed by David Beezer
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? Problem 5–1.19, *Transistion Mathematics*, Second Edition, Scott Foresman Addison Wesley, 1998. Solution [31]

M40 Contributed by Robert Beezer
Solutions to the system in Example IS [24] are given as

$$(x_1,\, x_2,\, x_3,\, x_4) = (-1 - 2\beta_3 + 3\beta_4,\, 4 + \beta_3 - 2\beta_4,\, \beta_3,\, \beta_4)$$

Evaluate the three equations of the original system with these expressions in β_3 and β_4 and verify that each equation is true, no matter what values are chosen for β_3 and β_4.

M70 Contributed by Robert Beezer
We have seen in this section that systems of linear equations have limited possibilities for solution sets, and we will shortly prove Theorem PSSLS [58] 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.

$$\begin{aligned} x^2 - y^2 &= 1 \\ x^2 + y^2 &= 4 \end{aligned}$$

After solving this system of *non-linear* equations, replace the second equation in turn by $x^2 + 2x + y^2 = 3$, $x^2 + y^2 = 1$, $x^2 - x + y^2 = 0$, $4x^2 + 4y^2 = 1$ and solve each resulting system of two equations in two variables. Solution [31]

T20 Contributed by Robert Beezer
Explain why the second equation operation in Definition EO [19] requires that the

scalar be nonzero, while this prohibition on the scalar is lifted in the third operation. Solution [31]

T10 Contributed by Robert Beezer
Technique D [15] asks you to formulate a definition of what it means for an integer 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. Solution [31]

Subsection SOL
Solutions

M30 Exercise [29] Contributed by Robert Beezer

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).

M70 Exercise [29] Contributed by Robert Beezer

The equation $x^2 - y^2 = 1$ has a solution set by itself that has the shape of a hyperbola when plotted. The five different second equations have solution sets that are circles when plotted individually. Where the hyperbola and circle intersect are the solutions to the system of two equations. As the size and location of the circle varies, the number of intersections varies from four to none (in the order given). Sketching the relevant equations would be instructive, as was discussed in Example STNE [15].

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

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

T10 Exercise [30] Contributed by Robert Beezer

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 $6 = 3 \times 2$, while 11 is odd since $11 = 5 \times 2 + 1$.

T20 Exercise [29] Contributed by Robert Beezer

Definition EO [19] is engineered to make Theorem EOPSS [20] 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 [20] 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.

Definition M
Matrix
An $m \times n$ **matrix** is a rectangular layout of numbers from $\mathbb{C}$ having m rows and n columns. △

Notation MN
Matrix Notation
We will use upper-case Latin letters from the start of the alphabet (A, B, $C, \dots$) 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). ⊠

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. ⊚

A calculator or computer language can be a convenient way to perform calculations with matrices. But first you have to enter the matrix. Here's how it is done on various computing platforms.

Computation Note ME.MMA
Matrix Entry (Mathematica)
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

$$\texttt{a} = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}, \{9, 10, 11, 12\}\}$$

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 `a//MatrixForm` , 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 it in the first place. ⊕

Computation Note ME.TI86
Matrix Entry (TI-86)
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,

$$[[1,2,3,4]\ [5,6,7,8]\ [9,10,11,12]] \rightarrow \mathtt{A}$$

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 ME.TI83
Matrix Entry (TI-83)
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, lets 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` . ⊕

Definition AM
Augmented Matrix

Suppose we have a system of m equations in the n variables x_1, x_2, $x_3, \ldots,$ x_n written as

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

then the **augmented matrix** of the system of equations is the $m \times (n+1)$ matrix

$$\begin{bmatrix}
a_{11} & a_{12} & a_{13} & \ldots & a_{1n} & b_1\\
a_{21} & a_{22} & a_{23} & \ldots & a_{2n} & b_2\\
a_{31} & a_{32} & a_{33} & \ldots & a_{3n} & b_3\\
\vdots & & & & & \\
a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} & b_m
\end{bmatrix}$$

△

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 [26].) 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 is the following system of 3 equations in 3 variables.

$$\begin{aligned}
x_1 - x_2 + 2x_3 &= 1\\
2x_1 + x_2 + x_3 &= 8\\
x_1 + x_2 &= 5
\end{aligned}$$

Here is its augmented matrix.

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

⊚

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 [19]) will preserve their solutions (Theorem EOPSS [20]). 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 entry in the same column 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 kind of shorthand to describe these 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.

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. △

Example TREM
Two row-equivalent matrices
The matrices

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

are row-equivalent as can be seen from

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

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 [48]), 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 [48]) 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. □

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 [20] 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 [23] as an exercise in using our new tools.

Example USR
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 [23] using equation operations.

$$\begin{aligned}x_1 + 2x_2 + 2x_3 &= 4\\ x_1 + 3x_2 + 3x_3 &= 5\\ 2x_1 + 6x_2 + 5x_3 &= 6\end{aligned}$$

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 [37] the system of equations below has the same solution set as the original system of equations.

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ x_2 + x_3 &= 1 \\ x_3 &= 4 \end{aligned}$$

Solving this "simpler" system is straightforward and is identical to the process in Example US [23]. ⊚

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. A row where every entry is zero is below any row containing 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 $i < s$, then $j < t$. △

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. Because we will make frequent reference to reduced row-echelon form, we make precise definitions of three terms.

Definition ZRM
Zero Row of a Matrix
A row of a matrix where every entry is zero is called a **zero row**. △

Definition LO
Leading Ones
For a matrix in reduced row-echelon form, the leftmost nonzero entry of any row that is not a zero row will be called a **leading 1**. △

Definition PC
Pivot Columns
For a matrix in reduced row-echelon form, a column containing a leading 1 will be called a **pivot column**. △

Example RREF
A matrix in reduced row-echelon form
The matrix C is in reduced row-echelon form.

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

Columns 1, 5, and 6 are pivot columns. ⊚

Example NRREF
A matrix not in reduced row-echelon form
The matrix D is not in reduced row-echelon form, as it fails each of the four requirements once.

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

⊚

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. Such is the case with our next theorem. ♢

Theorem REMEF
Row-Equivalent Matrix in Echelon Form
Suppose A is a matrix. Then there is a (unique!) 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. We will describe a process for converting A into B via row operations.

Set $i = 1$.

1. If $i = m + 1$, then stop converting the matrix.

2. Among all of the entries in rows i through m locate the leftmost nonzero entry (there may be several entries that tie for being leftmost). Denote the column of this entry by j. If this is not possible because all the entries are zero, then stop converting the matrix.

3. If the nonzero entry found in the preceding step is not in row i, swap rows so that row i has a nonzero entry in column j.

4. Use the second row operation to multiply row i by the reciprocal of the value in column j, thereby creating a leading 1 in row i at column j.

5. Use row i and the third row operation to convert every other entry in column j into a zero.

6. Increase i by one and return to step 1.

The result of this procedure is the matrix B. We need to establish that it has the requisite properties. First, the steps of the process only use row operations to convert the matrix, so A and B are row-equivalent.

It is a bit more work to be certain that B is in reduced row-echelon form. At the conclusion of the i-th trip through the steps, we claim the first i rows form a matrix in reduced row-echelon form, and the entries in rows $i+1$ through m in columns 1 through j are all zero. To see this, notice that

1. The definition of j insures that the entries of rows $i+1$ through m, in columns 1 through $j-1$ are all zero.

2. Row i has a leading nonzero entry equal to 1 by the result of step 4.

3. The employment of the leading 1 of row i in step 5 will make every element of column j zero in rows 1 through $i+1$, as well as in rows $i+1$ through m.

4. Rows 1 through $i-1$ are only affected by step 5. The zeros in columns 1 through $j-1$ of row i mean that none of the entries in columns 1 through $j-1$ for rows 1 through $i-1$ will change by the row operations employed in step 5.

5. Since columns 1 through j are are all zero for rows $i+1$ through m, any nonzero entry found on the next pass will be in a column to the right of column j, ensuring that the fourth condition of reduced row-echelon form is met.

6. If the procedure halts with $i = m+1$, then every row of B has a leading 1, and hence has no zero rows. If the procedure halts because step 2 fails to find a nonzero entry, then rows i through m are all zero rows, and they are all at the bottom of the matrix. ■

So now we can put it all together. Begin with a system of linear equations (Definition SSLE [16]), and represent it by its augmented matrix (Definition AM [34]). Use row operations (Definition RO [36]) to convert this matrix into reduced row-echelon form (Definition RREF [38]), using the procedure outlined in the proof of Theorem REMEF [39].

Theorem REMEF [39] also tells us we can always accomplish this, and that the result is row-equivalent (Definition REM [36]) to the original augmented matrix. Since the matrix in reduced-row echelon form has the same solution set, we can analyze it 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 [39] is unique. We could prove this result right now, but the proof will be much easier to state and understand a few sections from now when we have a few more definitions. However, the proof we will provide does not explicitly require any more *theorems* than we have right now, so we can, and will, make use of the uniqueness of B between now and then by citing Theorem RREFU [110]. You might want to jump forward now to read the statement of this important theorem and save studying its proof for later, once the rest of us get there.

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 them, circle them or write them in a different color. This device will prove very useful later and is a very good habit to start developing right now.

Example SAB
Solutions for Archetype B
Solve the following system of equations.

$$\begin{aligned} -7x_1 - 6x_2 - 12x_3 &= -33 \\ 5x_1 + 5x_2 + 7x_3 &= 24 \\ x_1 + 4x_3 &= 5 \end{aligned}$$

Form the augmented matrix for starters,

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

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

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

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

Now, with $i = 2$,

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

And finally, with $i = 3$,

$$\xrightarrow{\frac{5}{2}R_3} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & \boxed{1} & \frac{-13}{5} & \frac{-1}{5} \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{\frac{13}{5}R_3+R_2} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & \boxed{1} & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{-4R_3+R_1} \begin{bmatrix} \boxed{1} & 0 & 0 & -3 \\ 0 & \boxed{1} & 0 & 5 \\ 0 & 0 & \boxed{1} & 2 \end{bmatrix}$$

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. You might compare this example with the procedure we used in Example US [23]. ⊚

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

Example SAA
Solutions for Archetype A
Solve the following system of equations.

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 + x_2 + x_3 &= 8 \\ x_1 + x_2 &= 5 \end{aligned}$$

Form the augmented matrix for starters,

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

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

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

Now, with $i = 2$,

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

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 we can safely ignore it as we analyze the other two equations. These equations are,

$$\begin{aligned} x_1 + x_3 &= 3 \\ x_2 - x_3 &= 2. \end{aligned}$$

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.

$$\begin{aligned} x_1 &= 3 - x_3 \\ x_2 &= 2 + x_3 \end{aligned}$$

To write the solutions in set notation, we have

$$S = \{\, (3 - x_3, 2 + x_3,\, x_3) \mid x_3 \in \mathbb{C} \}$$

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 [24].⊚

Example SAE
Solutions for Archetype E
Solve the following system of equations.

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 2 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 3 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 2 \end{aligned}$$

Form the augmented matrix for starters,

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

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

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

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

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

Now, with $i = 2$,

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

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

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

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

And finally, with $i = 3$,

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

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

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

Lets 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, $\emptyset = \{\ \}$.

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 [41], Example SAA [42], Example SAE [43]) 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.

We will frequently use the term **row-reduce** as a verb. To row-reduce a matrix A will mean to apply row operations to A in order to find another matrix, B, that is in reduced row-echelon form and is row-equivalent to A. Theorem REMEF [39] tells us that this process will always be successful and Theorem RREFU [110] 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.C20 [**??**]).

Computation Note RR.MMA
Row Reduce (Mathematica)
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 RR.TI86
Row Reduce (TI-86)
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` . ⊕

Computation Note RR.TI83
Row Reduce (TI-83)
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` . ⊕

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.

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

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

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 0 \\ x_1 + 2x_2 + x_3 &= 3 \\ x_1 + 3x_2 + 3x_3 &= 7 \end{aligned}$$

Subsection EXC
Exercises

For problems C10–C12, find all solutions to the system of linear equations. Write the solutions as a set, using correct set notation.

C10 Contributed by Robert Beezer

$$\begin{aligned}
2x_1 - 3x_2 + x_3 + 7x_4 &= 14\\
2x_1 + 8x_2 - 4x_3 + 5x_4 &= -1\\
x_1 + 3x_2 - 3x_3 &= 4\\
-5x_1 + 2x_2 + 3x_3 + 4x_4 &= -19
\end{aligned}$$

Solution [49]

C11 Contributed by Robert Beezer

$$\begin{aligned}
3x_1 + 4x_2 - x_3 + 2x_4 &= 6\\
x_1 - 2x_2 + 3x_3 + x_4 &= 2\\
10x_2 - 10x_3 - x_4 &= 1
\end{aligned}$$

Solution [49]

C12 Contributed by Robert Beezer

$$\begin{aligned}
2x_1 + 4x_2 + 5x_3 + 7x_4 &= -26\\
x_1 + 2x_2 + x_3 - x_4 &= -4\\
-2x_1 - 4x_2 + x_3 + 11x_4 &= -10
\end{aligned}$$

Solution [49]

C30 Contributed by Robert Beezer

Row-reduce the matrix below without the aid of a calculator, indicating the row operations you are using at each step.

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

Solution [50]

M50 Contributed by Robert Beezer

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?
Solution [50]

T10 Contributed by Robert Beezer
Prove that each of the three row operations (Definition RO [36]) 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.

T11 Contributed by Robert Beezer
Suppose that A, B and C are $m \times n$ matrices. Use the definition of row-equivalence (Definition REM [36]) 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 [367].

Subsection SOL
Solutions

C10 Exercise [47] Contributed by Robert Beezer
The augmented matrix row-reduces to

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

and we see from the locations of the leading 1's that the system is consistent (Theorem RCLS [56]) and that $n-r=4-4=0$ and so the system has no free variables (Theorem CSRN [57]) and hence has a unique solution. This solution is $\{(1,\,-3,\,-4,\,1)\}$.

C11 Exercise [47] Contributed by Robert Beezer
The augmented matrix row-reduces to

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

and a leading 1 in the last column tells us that the system is inconsistent (Theorem RCLS [56]). So the solution set is $\emptyset = \{\}$.

C12 Exercise [47] Contributed by Robert Beezer
The augmented matrix row-reduces to

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

(Theorem RCLS [56]) and (Theorem CSRN [57]) tells us the system is consistent and the solution set can be described with $n-r=4-2=2$ free variables, namely x_2 and x_4. Solving for the dependent variables ($D=\{x_1,\,x_3\}$) the first and second equations represented in the row-reduced matrix yields,

$$\begin{aligned} x_1 &= 2 - 2x_2 + 4x_4 \\ x_3 &= -6 \qquad\;\; - 3x_4 \end{aligned}$$

As a set, we write this as

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

C30 Exercise [47] Contributed by Robert Beezer

$$\begin{bmatrix} 2 & 1 & 5 & 10 \\ 1 & -3 & -1 & -2 \\ 4 & -2 & 6 & 12 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 1 & -3 & -1 & -2 \\ 2 & 1 & 5 & 10 \\ 4 & -2 & 6 & 12 \end{bmatrix}$$

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

$$\xrightarrow{\frac{1}{7}R_2} \begin{bmatrix} 1 & -3 & -1 & -2 \\ 0 & 1 & 1 & 2 \\ 0 & 10 & 10 & 20 \end{bmatrix} \xrightarrow{3R_2+R_1} \begin{bmatrix} 1 & 0 & 2 & 4 \\ 0 & 1 & 1 & 2 \\ 0 & 10 & 10 & 20 \end{bmatrix}$$

$$\xrightarrow{-10R_2+R_3} \begin{bmatrix} \boxed{1} & 0 & 2 & 4 \\ 0 & \boxed{1} & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

M50 Exercise [47] Contributed by Robert Beezer

Let c, t, m, b denote the number of cars, trucks, motorcycles, and bicycles. Then the statements from the problem yield the equations:

$$\begin{aligned} c + t + m + b &= 66 \\ c - 4t &= 0 \\ 4c + 4t + 2m + 2b &= 252 \end{aligned}$$

The augmented matrix for this system is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 66 \\ 1 & -4 & 0 & 0 & 0 \\ 4 & 4 & 2 & 2 & 252 \end{bmatrix}$$

which row-reduces to

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 48 \\ 0 & \boxed{1} & 0 & 0 & 12 \\ 0 & 0 & \boxed{1} & 1 & 6 \end{bmatrix}$$

$c = 48$ is the first equation represented in the row-reduced matrix so there are 48 cars. $m + b = 6$ is the third equation represented in the row-reduced matrix 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 cannot have a negative number of motorcycles.

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.

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**. △

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, so we now initiate some notation that will help us talk about this form of a matrix.

Notation RREFA
Reduced Row-Echelon Form Analysis
Suppose that B is an $m \times n$ matrix that is in reduced row-echelon form. Let r equal the number of rows of B that are not zero rows. Each of these r rows then contains a leading 1, so let d_i equal the column number where row i's leading 1 is located. For columns without a leading 1, let f_i be the column number of the i-th column (reading from left to right) that does not contain a leading 1. Let

$$D = \{d_1,\, d_2,\, d_3,\, \ldots,\, d_r\} \qquad\qquad F = \{f_1,\, f_2,\, f_3,\, \ldots,\, f_{n-r}\}$$ ⊠

This notation can be a bit confusing, since we have subscripted variables that are in turn equal to subscripts used to index the matrix. However, many questions about matrices and systems of equations can be answered once we know r, D and F. An example may help.

Example RREFN
Reduced row-echelon form notation

For the 5×8 matrix

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

in reduced row-echelon form we have

$$\begin{array}{lllll} r = 4 & & & & \\ d_1 = 1 & d_2 = 3 & d_3 = 4 & d_4 = 7 & \\ f_1 = 2 & f_2 = 5 & f_3 = 6 & f_4 = 8 & f_5 = 9. \end{array}$$

Notice that the sets $D = \{d_1, d_2, d_3, d_4\} = \{1, 3, 4, 7\}$ and $F = \{f_1, f_2, f_3, f_4, f_5\} = \{2, 5, 6, 8, 9\}$ 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). ◎

Before proving some theorems about the possibilities for solution sets to systems of equations, lets 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 [510] is the system of $m = 4$ equations in $n = 7$ variables

$$\begin{aligned} x_1 + 4x_2 - x_4 + 7x_6 - 9x_7 &= 3 \\ 2x_1 + 8x_2 - x_3 + 3x_4 + 9x_5 - 13x_6 + 7x_7 &= 9 \\ 2x_3 - 3x_4 - 4x_5 + 12x_6 - 8x_7 &= 1 \\ -x_1 - 4x_2 + 2x_3 + 4x_4 + 8x_5 - 31x_6 + 37x_7 &= 4 \end{aligned}$$

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 [39]),

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

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.

$$\begin{aligned}
&(i = 1) \qquad && x_{d_1} = x_1 = 4 - 4x_2 - 2x_5 - x_6 + 3x_7\\
&(i = 2) \qquad && x_{d_2} = x_3 = 2 - x_5 + 3x_6 - 5x_7\\
&(i = 3) \qquad && x_{d_3} = x_4 = 1 - 2x_5 + 6x_6 - 6x_7
\end{aligned}$$

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 with elements that are 7-tuples, we write

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

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, mimicking the discussion in this example. We'll still be here when you get back. ⊚

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. So here's another proof technique.

Proof Technique SN
Set Notation
Sets are typically written inside of braces, as { }, and have two components. The first is a description of the 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 (:). Membership of an element in a set is denoted with the symbol $\in$.

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 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, 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.

$$E = \{x \in \mathbb{Z} \mid x \text{ is an even number}\} = \{x \in \mathbb{Z} \mid 2 \text{ divides } x \text{ evenly}\} = \{2k \mid k \in \mathbb{Z}\}.$$

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 ridiculous. $\diamond$

We mix our metaphors a bit when we call variables free versus dependent. Maybe we should call dependent variables "enslaved"? Here's the definition.

Definition IDV
Independent and Dependent Variables
Suppose A is the augmented matrix of a system of linear equations and B is a row-equivalent matrix in reduced row-echelon form. Suppose j is the number of a column of B that contains the leading 1 for some row, and it is not the last column. Then the variable j is **dependent**. A variable that is not dependent is called **independent** or **free**. $\triangle$

We can now use the values of m, n, r, and the independent and dependent variables to categorize the solutions sets to linear systems through a sequence of theorems. First the distinction between consistent and inconsistent systems, after two explanations of some proof techniques we will be using.

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 "A if and only if B," then it is true that "if A, then B" while it is also true that "if B, then A." 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 A and conclude B, then start over and assume B and conclude A. 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 A and B seem to be, the more pleasing it is to discover they are really equivalent. And if A describes a very mysterious solution or involves a tough computation, while B 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 $A \Rightarrow B$ is very easy, then proving $B \Rightarrow A$ 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. $\lozenge$

Proof Technique CP
Contrapositives

The **contrapositive** of an implication $A \Rightarrow B$ is the implication $\text{not}(B) \Rightarrow \text{not}(A)$, where "not" means the logical negation, or opposite. An implication is true if and only if its contrapositive is true. In symbols, $(A \Rightarrow B) \iff (\text{not}(B) \Rightarrow \text{not}(A))$ 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. ◊

Theorem RCLS
Recognizing Consistency of a Linear System
Suppose A is the augmented matrix of a system of linear equations with m equations in 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 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 [55]). Turning the implication around, and negating each portion, we arrive at the 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 i 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. Let $b_{i,n+1}$, $1 \leq i \leq r$ denote the entries of the last column of B for the first r rows. Employ our notation for columns of the reduced row-echelon form of a matrix (see Notation RREFA [51]) to B and set $x_{f_i} = 0$, $1 \leq i \leq n-r$ and then set $x_{d_i} = b_{i,n+1}$, $1 \leq i \leq r$. These values for the variables make the equations represented by the first r rows all true (convince yourself of this). Rows $r+1$ through m (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, so we can say it is consistent. ■

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.

Theorem ICRN
Inconsistent Systems, r and n
Suppose A is the augmented matrix of a system of linear equations with m 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 $D = \{1, 2, 3, \ldots, n, n+1\}$ and every column of B contains a leading 1. In particular, the entry of column $n+1$ for row $r = n+1$ is a leading 1. Theorem RCLS [56] then says that the system is inconsistent. ■

Theorem CSRN
Consistent Systems, r and n
Suppose A is the augmented matrix of a *consistent* system of linear equations with m equations in 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 the echelon layout of the leading 1's means that there are at most $n + 1$ leading 1's and therefore $r \leq n + 1$ for any system. We are assuming this system is consistent, so we know by Theorem ICRN [56] that $r \neq n + 1$. Together these two observations leave us with $r \leq n$.

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 [56]. By varying the values of the free variables suitably, we can demonstrate infinitely many solutions. ■

Theorem FVCS
Free Variables for Consistent Systems
Suppose A is the augmented matrix of a *consistent* system of linear equations with m 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. Then the solution set can be described with $n - r$ free variables. □

Proof Technique CV
Converses
The **converse** of the implication $A \Rightarrow B$ is the implication $B \Rightarrow A$. 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 [54]). 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 [57] has a tempting converse. Does this theorem say that if $r < n$, then the system is consistent? Definitely not, as Archetype E [491] has $r = 2$ and $n = 4$ but 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. ♢

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 [473] 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 [478] has $n = 3$ and $r = 3$. It can be seen to be consistent by the single sample solution given. Its solution set then has $n - r = 0$ free variables, and therefore will have just the single solution.

3. Archetype H [505] has $n = 2$ and $r = 3$. In this case, $r = n + 1$, so Theorem ICRN [56] says the system is inconsistent. We should not try to apply Theorem FVCS [57] to count free variables, since the theorem only applies to consistent systems. (What would happen if you did?)

4. Archetype E [491] has $n = 4$ and $r = 3$. However, by looking at the reduced row-echelon form of the augmented matrix, we find a leading 1 in row 3, column 4. By Theorem RCLS [56] we recognize the system is then inconsistent. (Why doesn't this example contradict Theorem ICRN [56]?) ⊚

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. Notice that this theorem was presaged first by Example TTS [17] and further foreshadowed by other examples.

Theorem PSSLS
Possible Solution Sets for Linear Systems
A simultaneous system of linear equations has no solutions, a unique solution or infinitely many solutions. □

Proof By definition, a system is either inconsistent or consistent. 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 [57]. ■

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

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

A consistent system with free variables will have an infinite number of solutions, as given by Theorem CSRN [57]. ■

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,

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 8 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= -12 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 4 \end{aligned}$$

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 [59] and conclude that the system has infinitely many solutions. ⊚

These theorems give us the procedures and implications that allow us to completely solve any simultaneous 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 [39].

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 [56]. 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.

Computation Note LS.MMA
Linear Solve (Mathematica)

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 [56], 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 [473] is

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 + x_2 + x_3 &= 8 \\ x_1 + x_2 &= 5 \end{aligned}$$

To ask *Mathematica* for a solution, enter

```
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. ⊕

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 do we recognize when a system of linear equations is inconsistent?
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

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** Contributed by Robert Beezer

A homogeneous system of 8 equations in 8 variables. Solution [65]

M51 Contributed by Robert Beezer

A consistent system of 8 equations in 6 variables. Solution [65]

M52 Contributed by Robert Beezer

A consistent system of 6 equations in 8 variables. Solution [65]

Subsection SOL
Solutions

M50 Exercise [63] Contributed by Robert Beezer
Since the system is homogeneous, we know it has the trivial solution (Theorem HSC [68]). 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 [58]). See Archetype A [473] and Archetype B [478] to understand the possibilities.

M51 Exercise [63] Contributed by Robert Beezer
Consistent means there is at least one solution (Definition CS [51]). It will have either a unique solution or infinitely many solutions (Theorem PSSLS [58]).

M52 Exercise [63] Contributed by Robert Beezer
With 6 rows in the augmented matrix, the row-reduced version will have $r \leq 6$. Since the system is consistent, apply Theorem CSRN [57] to see that $n-r \geq 2$ implies infinitely many solutions.

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 is **homogeneous** if each equation has a 0 for its constant term. Such a system then has the form,

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= 0\\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= 0\\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= 0\\
&\vdots\\
a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= 0
\end{aligned}$$

△

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, we can convert the original system of equations into the homogeneous system,

$$\begin{aligned}
2x_1 - 3x_2 + x_3 - 6x_4 &= 0\\
4x_1 + x_2 + 2x_3 + 9x_4 &= 0\\
3x_1 + x_2 + x_3 + 8x_4 &= 0
\end{aligned}$$

Can you quickly find a solution to this system without row-reducing the augmented matrix? ⊚

As you might have discovered by studying Example AHSAC [67], 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 it is consistent. □

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 $x_1 = 0$, $x_2 = 0, \ldots, x_n = 0$ 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,

$$\begin{aligned} -11x_1 + 2x_2 - 14x_3 &= 0 \\ 23x_1 - 6x_2 + 33x_3 &= 0 \\ 14x_1 - 2x_2 + 17x_3 &= 0 \end{aligned}$$

whose augmented matrix row-reduces to

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \end{bmatrix}$$

By Theorem HSC [68], 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 [473] can be converted to the homogeneous system,

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 0 \\ 2x_1 + x_2 + x_3 &= 0 \\ x_1 + x_2 \qquad &= 0 \end{aligned}$$

whose augmented matrix row-reduces to

$$\begin{bmatrix} \boxed{1} & 0 & 1 & 0 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem HSC [68], the system is consistent, and so the computation $n-r = 3-2 = 1$ means the solution set contains one free variable by Theorem FVCS [57], and hence has infinitely many solutions. We can describe this solution set using the free variable x_3,

$$S = \{(x_1,\,x_2,\,x_3) \mid x_1 = -x_3,\, x_2 = x_3\} = \{(-x_3,\,x_3,\,x_3) \mid x_3 \in \mathbb{C}\}$$

Geometrically, these are points in three dimensions that lie on a line through the origin.⊚

Example HISAD
Homogeneous, infinite solutions, Archetype D
Archetype D [487] (and identically, Archetype E [491]) can be converted to the homogeneous system,

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 0 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 0 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 0 \end{aligned}$$

whose augmented matrix row-reduces to

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

By Theorem HSC [68], the system is consistent, and so the computation $n-r = 4-2 = 2$ means the solution set contains two free variables by Theorem FVCS [57], and hence has infinitely many solutions. We can describe this solution set using the free variables x_3 and x_4,

$$\begin{aligned} S &= \{(x_1,\,x_2,\,x_3,\,x_4) \mid x_1 = -3x_3 + 2x_4,\, x_2 = -x_3 + 3x_4\} \\ &= \{(-3x_3 + 2x_4,\, -x_3 + 3x_4,\, x_3,\, x_4) \mid x_3,\, x_4 \in \mathbb{C}\} \end{aligned}$$

⊚

After working through these examples, you might perform the same computations for the slightly larger example, Archetype J [515].

Example HISAD [69] 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 [68] says it is consistent. Then the hypothesis that $n > m$, together with Theorem CMVEI [59], gives infinitely many solutions. ■

Example HUSAB [68] and Example HISAA [68] 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 [69] tells us that there is only one possibility for a homogeneous system).

Subsection MVNSE
Matrix and Vector Notation for Systems of Equations

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. This observation might suffice as a first explanation of the reason for some of the following definitions.

Definition CV
Column Vector
A **column vector** of **size** m is an ordered list of m numbers, which is written vertically, in order from top to bottom. At times, we will refer to a column vector as simply a **vector**. △

Notation VN
Vector (u)
Column vectors will be written in bold, usually with lower case letters $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$. 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 $\underset{\sim}{u}$. ⊠

Definition ZV
Zero Vector
The **zero vector** of size m is the column vector of size m where each entry is the number zero,

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

Notation ZVN
Zero Vector (0)
The zero vector will be written as $\mathbf{0}$. ⊠

Definition CM
Coefficient Matrix
For a system of linear equations,

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

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,

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

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 SV
Solution Vector
For a system of linear equations,

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

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

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

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.

Notation AMN
Augmented Matrix ($[A\,|\,\mathbf{b}]$)
With these definitions, we will write the augmented matrix of system of linear equations in the form $[A\,|\,\mathbf{b}]$ in order to identify and distinguish the coefficients and the constants.⊠

Notation LSN
Linear System ($\mathcal{LS}(A,\,\mathbf{b})$)
We will write $\mathcal{LS}(A,\,\mathbf{b})$ to denote the system of linear equations with A as a coefficient matrix and $\mathbf{b}$ as the vector of constants. ⊠

Example NSLE
Notation for systems of linear equations
The system of linear equations

$$\begin{aligned} 2x_1 + 4x_2 - 3x_3 + 5x_4 + x_5 &= 9 \\ 3x_1 + x_2 + \quad + x_4 - 3x_5 &= 0 \\ -2x_1 + 7x_2 - 5x_3 + 2x_4 + 2x_5 &= -3 \end{aligned}$$

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 referenced as $\mathcal{LS}(A, \mathbf{b})$. ⊚

With these definitions and notation a homogeneous system will be indicated by $\mathcal{LS}(A, \mathbf{0})$. Its augmented matrix will be $[A \,|\, \mathbf{0}]$, which when converted to reduced row-echelon form will still have the final column of zeros. So in this case, we may be as likely to just reference only the coefficient matrix.

Subsection NSM
Null Space of a Matrix

The set of solutions to a homogeneous system (which by Theorem HSC [68] 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 $\mathcal{LS}(A, \mathbf{0})$. △

In the Archetypes (Chapter A [469]) 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 [510] 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 [510].

$$\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ -5 \\ -6 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} -4 \\ 1 \\ -3 \\ -2 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

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. ⊚

Subsection READ
Reading Questions

1. What is *always* true of the solution set for a homogenous system of equations?
2. Suppose a homogenous sytem 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.

Section NSM
NonSingular Matrices

In this section we specialize to systems with equal numbers of equations and variables, which will prove to be a case of special interest.

Subsection NSM
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. Now would be a good time to review the discussion about speaking and writing mathematics in Technique L [26].

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. And suppose the homogeneous linear system of equations $\mathcal{LS}(A,\, \mathbf{0})$ has *only* the trivial solution. Then we say that A is a **nonsingular** matrix. Otherwise we say A is a **singular** matrix. △

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 (homogenous, 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 homogenous system of equations.

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 [68] shows that the coefficient matrix derived from Archetype A [473], 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 NS
A nonsingular matrix, Archetype B
Example HUSAB [68] shows that the coefficient matrix derived from Archetype B [478], 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 solution. ⊚

Notice that we will not discuss Example HISAD [69] 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**, $I_m = (a_{ij})$ has $a_{ij} = 1$ whenever $i = j$, and $a_{ij} = 0$ whenever $i \neq j$. △

Example IM
An identity matrix
The 4×4 identity matrix is

$$I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$ ⊚

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 NSRRI
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. □

Proof ($\Leftarrow$) Suppose B is the identity matrix. When the augmented matrix $[A \mid \mathbf{0}]$ is row-reduced, 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 [57] 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.

($\Rightarrow$) We will prove the contrapositive. Suppose B is not the identity matrix. When the augmented matrix $[A \mid \mathbf{0}]$ is row-reduced, the result is $[B \mid \mathbf{0}]$. The number of nonzero rows is less than the number of variables for the system of equations, so Theorem FVCS [57] gives $n - r > 0$ free variables. Thus, the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has infinitely many solutions, so the system has more solutions than just the trivial solution. Thus the matrix is not nonsingular, the desired conclusion. ■

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 [473] is

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

which when row-reduced becomes the row-equivalent matrix

$$B = \begin{bmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{1} & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since this matrix is not the 3×3 identity matrix, Theorem NSRRI [76] tells us that A is a singular matrix. ⊚

Example NSRR
NonSingular matrix, row-reduced
The coefficient matrix for Archetype B [478] is

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

which when row-reduced becomes the row-equivalent matrix

$$B = \begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{bmatrix}.$$

Since this matrix is the 3×3 identity matrix, Theorem NSRRI [76] tells us that A is a nonsingular matrix. ⊚

Example NSS
Null space of a singular matrix
Given the coefficient matrix from Archetype A [473],

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

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 [68] as

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} -x_3 \\ x_3 \\ x_3 \end{bmatrix} \,\middle|\, x_3 \in \mathbb{C} \right\}$$ ⊚

Example NSNS
Null space of a nonsingular matrix
Given the coefficient matrix from Archetype B,

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

the homogeneous system $\mathcal{LS}(A,\,\mathbf{0})$ has a solution set constructed in Example HUSAB [68] 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.

Theorem NSTNS
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 trivial solution to the system $\mathcal{LS}(A,\,\mathbf{0})$, i.e. $\mathcal{N}(A) = \{\mathbf{0}\}$. □

Proof The null space of a square *matrix*, A, is 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 of half of this theorem. ■

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, or the conclusion that the two allegedly different objects really are equal. ◊

The next theorem pulls a lot of ideas together. It tells us that we can learn a lot about solutions to a system of linear equations with a square coefficient matrix by examining a similar homogeneous system.

Theorem NSMUS
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 [75]). 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.

If the first half of the proof seemed easy, perhaps we'll have to work a bit harder to get the implication in the opposite direction. We provide two different proofs for the second half. The first is suggested by Asa Scherer and relies on the uniqueness of the reduced row-echelon form of a matrix (Theorem RREFU [110]), a result that we could have proven earlier, but we have decided to delay until later. The second proof is lengthier and more involved, but does not rely on the uniqueness of the reduced row-echelon form of a matrix, a result we have not proven yet. It is also a good example of the types of proofs we will encounter throughout the course.

($\Rightarrow$, Round 1) We assume that A is nonsingular, so we know there is a sequence of row operations that will convert A into the identity matrix I_n (Theorem NSRRI [76]). Form the augmented matrix $A'=[\,A\,|\,b]$ and apply this same sequence of row operations to A'. The result will be the matrix $B'=[\,I_n\,|\,c]$, which is in reduced row-echelon form. It should be clear that $\mathbf{c}$ is a solution to $\mathcal{LS}(A,\,b)$. Furthermore, since B' is unique (Theorem RREFU [110]), the vector $\mathbf{c}$ must be unique, and therefore is a unique solution of $\mathcal{LS}(A,\,b)$.

($\Rightarrow$, Round 2) We will assume A is nonsingular, and try to solve the system $\mathcal{LS}(A,\,\mathbf{b})$ without making any assumptions about $\mathbf{b}$. 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,

$$\begin{aligned}
a_{11}x_1+a_{12}x_2+a_{13}x_3+\cdots+a_{1n}x_n&=b_1\\
a_{21}x_1+a_{22}x_2+a_{23}x_3+\cdots+a_{2n}x_n&=b_2\\
a_{31}x_1+a_{32}x_2+a_{33}x_3+\cdots+a_{3n}x_n&=b_3\\
&\vdots\\
a_{n1}x_1+a_{n2}x_2+a_{n3}x_3+\cdots+a_{nn}x_n&=b_n
\end{aligned}\qquad (*)$$

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,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n - b_1y &= 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n - b_2y &= 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n - b_3y &= 0 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \cdots + a_{nn}x_n - b_ny &= 0 \end{aligned} \qquad (**)$$

Since this is a homogeneous system with more variables than equations ($m = n+1 > n$), Theorem HMVEI [69] 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

$$x_1 = c_1 \qquad x_2 = c_2 \qquad x_3 = c_3 \qquad \ldots \qquad x_n = c_n \qquad y = c_{n+1}$$

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. 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,

$$a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \cdots + a_{in}c_n - b_i(0) = 0$$

which becomes

$$a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \cdots + a_{in}c_n = 0$$

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

$$x_1 = \frac{c_1}{c_{n+1}} \qquad x_2 = \frac{c_2}{c_{n+1}} \qquad x_3 = \frac{c_3}{c_{n+1}} \qquad \ldots \qquad x_n = \frac{c_n}{c_{n+1}}$$

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 $(**)$,

$$\begin{aligned}
&a_{i1}\frac{c_1}{c_{n+1}} + a_{i2}\frac{c_2}{c_{n+1}} + a_{i3}\frac{c_3}{c_{n+1}} + \cdots + a_{in}\frac{c_n}{c_{n+1}}\\
&= \frac{1}{c_{n+1}}\left(a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \cdots + a_{in}c_n\right)\\
&= \frac{1}{c_{n+1}}\left(a_{i1}c_1 + a_{i2}c_2 + a_{i3}c_3 + \cdots + a_{in}c_n - b_ic_{n+1}\right) + b_i\\
&= \frac{1}{c_{n+1}}\left(0\right) + b_i\\
&= b_i
\end{aligned}$$

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,

$$\begin{array}{lllll}
x_1 = d_1 & x_2 = d_2 & x_3 = d_3 & \dots & x_n = d_n\\
x_1 = e_1 & x_2 = e_2 & x_3 = e_3 & \dots & x_n = e_n
\end{array}$$

Then,

$$\begin{aligned}
&(a_{i1}(d_1 - e_1) + a_{i2}(d_2 - e_2) + a_{i3}(d_3 - e_3) + \cdots + a_{in}(d_n - e_n))\\
&= (a_{i1}d_1 + a_{i2}d_2 + a_{i3}d_3 + \cdots + a_{in}d_n) - (a_{i1}e_1 + a_{i2}e_2 + a_{i3}e_3 + \cdots + a_{in}e_n)\\
&= b_i - b_i\\
&= 0
\end{aligned}$$

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. ■

This important theorem deserves several comments. First, notice that the proposed solution ($x_i = \frac{c_i}{c_{n+1}}$) appeared in the Round 2 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.

Second, 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 [211]).

Finally, 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 [206]).

Proof Technique ME
Multiple Equivalences
A very specialized form of a theorem begins with the statement "The following are equivalent..." and then follows 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,... Z. To prove the entire theorem, we can prove $A \Rightarrow B$, $B \Rightarrow C$, $C \Rightarrow D$,..., $Y \Rightarrow Z$ 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. ♢

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 have all of the opposite properties.

Theorem NSME1
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 trivial solution, $\mathcal{N}(A) = \{\mathbf{0}\}$.
4. The linear system $\mathcal{LS}(A, \mathbf{b})$ has a unique solution for every possible choice of $\mathbf{b}$. □

Proof That A is nonsingular is equivalent to each of the subsequent statements by, in turn, Theorem NSRRI [76], Theorem NSTNS [78] and Theorem NSMUS [79]. So the statement of this theorem is just a convenient way to organize all these results. ■

Subsection READ
Reading Questions

1. What is the definition of a nonsingular matrix?

2. What is the easiest way to recognize a nonsingular matrix?

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

C30 Contributed by Robert Beezer
Is the matrix below singular or nonsingular? Why?

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

Solution [87]

Subsection SOL
Solutions

C30 Exercise [85] Contributed by Robert Beezer

The matrix row-reduces to

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

which is the 4×4 identity matrix. By Theorem NSRRI [76] the original matrix must be nonsingular.

V: Vectors

Section VO
Vector Operations

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. Initially we will depart from our study of systems of linear equations, but in Section LC [97] we will forge a connection between linear combinations and systems of linear combinations in Theorem SLSLC [101]. This connection will allow us to understand systems of linear equations at a higher level, while consequently discussing them less frequently.

In the current 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 a matrix with just one column (Definition CV [70]). The collection of all possible vectors of a fixed size is a commonly used set, so we start with its definition.

Definition VSCM
Vector Space $\mathbb{C}^m$
The vector space $\mathbb{C}^m$ is the set of all column vectors (Definition CV [70]) of size m with entries from the set of complex numbers, $\mathbb{C}$. △

When this set is defined using only entries 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 a matrix with a single column. It could simply be an ordered list of numbers, and written like $(2, 3, -1, 6)$. 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.

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 a 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
The vectors

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_m \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix}$$

are **equal**, written $\mathbf{u} = \mathbf{v}$ provided that $u_i = v_i$ for all $1 \leq i \leq m$. △

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.

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. Earlier, in Technique SE [21] we discussed at some length 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, lets 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 simultaneous linear equations in Archetype B [478],

$$\begin{aligned} -7x_1 - 6x_2 - 12x_3 &= -33 \\ 5x_1 + 5x_2 + 7x_3 &= 24 \\ x_1 + 4x_3 &= 5 \end{aligned}$$

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 [90], 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. ⊚

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
Given the vectors

$$\mathbf{u} = \begin{bmatrix}u_1\\ u_2\\ u_3\\ \vdots\\ u_m\end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix}v_1\\ v_2\\ v_3\\ \vdots\\ v_m\end{bmatrix}$$

the **sum** of $\mathbf{u}$ and $\mathbf{v}$ is the vector

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix}u_1\\ u_2\\ u_3\\ \vdots\\ u_m\end{bmatrix} + \begin{bmatrix}v_1\\ v_2\\ v_3\\ \vdots\\ v_m\end{bmatrix} = \begin{bmatrix}u_1 + v_1\\ u_2 + v_2\\ u_3 + v_3\\ \vdots\\ u_m + v_m\end{bmatrix}.$$ △

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 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

$$\mathbf{u} = \begin{bmatrix}2\\ -3\\ 4\\ 2\end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix}-1\\ 5\\ 2\\ -7\end{bmatrix}$$

then

$$\mathbf{u}+\mathbf{v}=\begin{bmatrix}2\\-3\\4\\2\end{bmatrix}+\begin{bmatrix}-1\\5\\2\\-7\end{bmatrix}=\begin{bmatrix}2+(-1)\\-3+5\\4+2\\2+(-7)\end{bmatrix}=\begin{bmatrix}1\\2\\6\\-5\end{bmatrix}.$$ ⊚

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
Given the vector

$$\mathbf{u}=\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_m\end{bmatrix}$$

and the scalar $\alpha \in \mathbb{C}$, the **scalar multiple** of $\mathbf{u}$ by α is

$$\alpha\mathbf{u}=\alpha\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_m\end{bmatrix}=\begin{bmatrix}\alpha u_1\\\alpha u_2\\\alpha u_3\\\vdots\\\alpha u_m\end{bmatrix}.$$ △

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, ...) and write vectors in bold Latin letters from the end of the alphabet ($\mathbf{u}$, $\mathbf{v}$, ...), 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.

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

$$\alpha\mathbf{u} = 6 \begin{bmatrix} 3 \\ 1 \\ -2 \\ 4 \\ -1 \end{bmatrix} = \begin{bmatrix} 6(3) \\ 6(1) \\ 6(-2) \\ 6(4) \\ 6(-1 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \\ -12 \\ 24 \\ -6 \end{bmatrix}.$$

⊚

It is usually straightforward to effect these computations with a calculator or program.

Computation Note VLC.MMA
Vector Linear Combinations (Mathematica)
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

$$\texttt{v} = \{\texttt{1, 2, 3, 4}\}$$

Vector addition and scalar multiplication are then very natural. If `u` and `v` are two lists of equal length, then

$$\texttt{2u} + (-\texttt{3})\texttt{v}$$

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 VLC.TI86
Vector Linear Combinations (TI-86)
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`),

$$\texttt{[1, 2, 3, 4]} \rightarrow \texttt{v}$$

Vector addition and scalar multiplication are then very natural. If `u` and `v` are two vectors of equal size, then

$$\texttt{2} * \texttt{u} + (-\texttt{3}) * \texttt{v}$$

will compute the correct vector and display the result as a vector. ⊕

Computation Note VLC.TI83
Vector Linear Combinations (TI-83)
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` . ⊕

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 VSPCM
Vector Space Properties of $\mathbb{C}^m$
Suppose that $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{C}^m$ and α and β are scalars. Then

1. $\mathbf{u} + \mathbf{v} \in \mathbb{C}^m$ (Additive closure)

2. $\alpha\mathbf{u} \in \mathbb{C}^m$ (Scalar closure)

3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (Commutativity)

4. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (Associativity of vector addition)

5. There is a vector, $\mathbf{0}$, called the **zero vector**, such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$ for all $\mathbf{u}\in\mathbb{C}^m$. (Additive identity)

6. For each vector $\mathbf{u}\in\mathbb{C}^m$, there exists a vector $-\mathbf{u}\in\mathbb{C}^m$ so that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$. (Additive inverses)

7. $\alpha(\beta\mathbf{u})=(\alpha\beta)\mathbf{u}$ (Associativity of scalar multiplication)

8. $\alpha(\mathbf{u}+\mathbf{v})=\alpha\mathbf{u}+\alpha\mathbf{v}$ (Distributivity across vector addition)

9. $(\alpha+\beta)\mathbf{u}=\alpha\mathbf{u}+\beta\mathbf{u}$ (Distributivity across addition)

10. $1\mathbf{u}=\mathbf{u}$ (Scalar multiplication with 1) □

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.

$$\begin{aligned}(\alpha+\beta)\mathbf{u}=(\alpha+\beta)\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_m\end{bmatrix}&=\begin{bmatrix}(\alpha+\beta)u_1\\(\alpha+\beta)u_2\\(\alpha+\beta)u_3\\\vdots\\(\alpha+\beta)u_m\end{bmatrix}=\begin{bmatrix}\alpha u_1+\beta u_1\\\alpha u_2+\beta u_2\\\alpha u_3+\beta u_3\\\vdots\\\alpha u_m+\beta u_m\end{bmatrix}\\&=\begin{bmatrix}\alpha u_1\\\alpha u_2\\\alpha u_3\\\vdots\\\alpha u_m\end{bmatrix}+\begin{bmatrix}\beta u_1\\\beta u_2\\\beta u_3\\\vdots\\\beta u_m\end{bmatrix}=\alpha\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_m\end{bmatrix}+\beta\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_m\end{bmatrix}\\&=\alpha\mathbf{u}+\beta\mathbf{u}\end{aligned}$$

■

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

$$-\mathbf{u}=\begin{bmatrix}-u_1\\-u_2\\-u_3\\\vdots\\-u_m\end{bmatrix}=(-1)\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_m\end{bmatrix}=(-1)\mathbf{u}$$

We will see later how to derive this property as a *consequence* of several of the ten properties listed in Theorem VSPCM [94].

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

$$2\begin{bmatrix}1\\5\\0\end{bmatrix}+(-3)\begin{bmatrix}7\\6\\5\end{bmatrix}$$

Section LC
Linear Combinations

Subsection LC
Linear Combinations

In Section VO [89] 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.

Definition LCCV
Linear Combination of Column Vectors
Given n vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\ldots$, $\mathbf{u}_n$ and n scalars α_1, α_2, α_3, $\ldots$, α_n, their **linear combination** is the vector

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n. \qquad \triangle$$

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 brand-new 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

$$\alpha_1 = 3 \qquad \alpha_2 = -4 \qquad \alpha_3 = 2 \qquad \alpha_4 = -1$$

and

$$\mathbf{u}_1 = \begin{bmatrix} 2\\ 4\\ -3\\ 1\\ 2\\ 9 \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} 6\\ 3\\ 0\\ -2\\ 1\\ 4 \end{bmatrix} \qquad \mathbf{u}_3 = \begin{bmatrix} -5\\ 2\\ 1\\ 1\\ -3\\ 0 \end{bmatrix} \qquad \mathbf{u}_4 = \begin{bmatrix} 3\\ 2\\ -5\\ 7\\ 1\\ 3 \end{bmatrix}$$

then their linear combination is

$$\alpha_1\mathbf{u_1} + \alpha_2\mathbf{u_2} + \alpha_3\mathbf{u_3} + \alpha_4\mathbf{u_4} = (1)\begin{bmatrix}2\\4\\-3\\1\\2\\9\end{bmatrix} + (-4)\begin{bmatrix}6\\3\\0\\-2\\1\\4\end{bmatrix} + (2)\begin{bmatrix}-5\\2\\1\\1\\-3\\0\end{bmatrix} + (-1)\begin{bmatrix}3\\2\\-5\\7\\1\\3\end{bmatrix}$$
$$= \begin{bmatrix}2\\4\\-3\\1\\2\\9\end{bmatrix} + \begin{bmatrix}-24\\-12\\0\\8\\-4\\-16\end{bmatrix} + \begin{bmatrix}-10\\4\\2\\2\\-6\\0\end{bmatrix} + \begin{bmatrix}-3\\-2\\5\\-7\\-1\\-3\end{bmatrix} = \begin{bmatrix}-35\\-6\\4\\4\\-9\\-8\end{bmatrix}.$$

A different linear combination, of the same set of vectors, can be formed with different scalars. Take

$$\beta_1 = 3 \qquad \beta_2 = 0 \qquad \beta_3 = 5 \qquad \beta_4 = -6$$

and form the linear combination

$$\beta_1\mathbf{u_1} + \beta_2\mathbf{u_2} + \beta_3\mathbf{u_3} + \beta_4\mathbf{u_4} = (3)\begin{bmatrix}2\\4\\-3\\1\\2\\9\end{bmatrix} + (0)\begin{bmatrix}6\\3\\0\\-2\\1\\4\end{bmatrix} + (5)\begin{bmatrix}-5\\2\\1\\1\\-3\\0\end{bmatrix} + (-6)\begin{bmatrix}3\\2\\-5\\7\\1\\3\end{bmatrix}$$
$$= \begin{bmatrix}6\\12\\-9\\3\\6\\27\end{bmatrix} + \begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix} + \begin{bmatrix}-25\\10\\5\\5\\-15\\0\end{bmatrix} + \begin{bmatrix}-16\\-2\\5\\-7\\-1\\-3\end{bmatrix} = \begin{bmatrix}-35\\20\\1\\1\\-10\\24\end{bmatrix}.$$

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. ⊚

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 building blocks. ♢

Example ABLC
Archetype B as a linear combination
In this example we will rewrite Archetype B [478] in the language of vectors, vector equality and linear combinations. In Example VESE [90] we wrote the simultaneous 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 [478], 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 NSRRI [76] to determine that the coefficient matrix is nonsingular. Then Theorem NSMUS [79] tells us that the system of equations has a unique solution. This solution is

$$x_1 = -3 \qquad\qquad x_2 = 5 \qquad\qquad x_3 = 2.$$

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. ⊚

With any discussion of Archetype A [473] or Archetype B [478] we should be sure to contrast with the other.

Example AALC
Archetype A as a linear combination
As a vector equality, Archetype A [473] 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 [473] leads to the conclusion that the system is consistent and has free variables, hence infinitely many solutions. So for example, the two solutions

$$\begin{aligned}x_1&=2 & x_2&=3 & x_3&=1\\x_1&=3 & x_2&=2 & x_3&=0\end{aligned}$$

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 [473]. 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

$$x_1 = -1 \qquad x_2 = 1 \qquad x_3 = 1$$

is a nontrivial solution to the homogeneous system of equations with the coefficient matrix for the original system in Archetype A [473]. In particular, this demonstrates that this coefficient matrix is singular. ⊚

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 $\mathbf{x} = \begin{bmatrix}\alpha_1\\\alpha_2\\\alpha_3\\\vdots\\\alpha_n\end{bmatrix}$ is a solution to the linear system of equations $\mathcal{LS}(A, \mathbf{b})$ if and only if

$$\alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2 + \alpha_3\mathbf{A}_3 + \cdots + \alpha_n\mathbf{A}_n = \mathbf{b}$$ □

Proof Write the system of equations $\mathcal{LS}(A, b)$ as

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

Now use vector equality (Definition CVE [90]) to replace the m simultaneous equalities by one vector equality,

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

Use vector addition (Definition CVA [91]) to rewrite,

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

And finally, use the definition of vector scalar multiplication Definition CVSM [92],

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

and use notation for the various column vectors,

$$x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 + \cdots + x_n\mathbf{A}_n = \mathbf{b}.$$

Each of the expressions above is just a rewrite of another one. So if we begin with a solution to the system of equations, substituting its values into the original system will make the equations simultaneously true. But then these same values will also make the final expression with the linear combination true. Reversing the argument, and employing the equations in reverse, will give the other half of the proof. ■

In other words, this theorem tells us that solutions to systems of equations are linear combinations of the column vectors of the coefficient matrix ($\mathbf{A}_i$) which yield the constant vector $\mathbf{b}$. 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 realize the vector $\mathbf{b}$ 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.

Subsection VFSS
Vector Form of Solution Sets

We have recently begun writing solutions to systems of equations as column vectors. For example Archetype B [478] 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 an example 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 [487] is a linear system of 3 equations in 4 variables. Row-reducing the augmented matrix yields

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

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 develop a linear combination that expresses a typical solution, in three steps.

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.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} + x_3 \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} + x_4 \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

Step 2. For each free variable, use 0's and 1's to ensure equality for the corresponding entry of the the vectors.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \\ \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \\ \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} \\ \\ 0 \\ 1 \end{bmatrix}$$

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.

$$x_1 = 4 - 3x_3 + 2x_4 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ \\ 0 \\ 1 \end{bmatrix}$$

$$x_2 = 0 - 1x_3 + 3x_4 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

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

$$x_3 = 2,\ x_4 = -5 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (2) \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + (-5) \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -12 \\ -17 \\ 2 \\ -5 \end{bmatrix}$$

or,

$$x_3 = 1,\ x_4 = 3 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} + (1) \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + (3) \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 8 \\ 1 \\ 3 \end{bmatrix}.$$

You'll find the second solution listed in the write-up for Archetype D [487], and you might check the first solution by substituting it back into the original equations.

While this form is useful for quickly creating solutions, its 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. ⊚

We'll now formalize the last example 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. Denote the vector of variables as $\mathbf{x} = (x_i)$. Let $B = (b_{ij})$ be a row-equivalent $m\times(n+1)$ matrix in reduced row-echelon form. Suppose that B has r nonzero rows, columns without leading 1's having 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 $\mathbf{c} = (c_i)$, $\mathbf{u}_j = (u_{ij})$, $1 \leq j \leq n-r$ of size n by

$$c_i = \begin{cases} 0 & \text{if } i \in F \\ b_{k,n+1} & \text{if } i \in D,\ i = d_k \end{cases}$$

$$u_{ij} = \begin{cases} 1 & \text{if } i \in F,\ i = f_j \\ 0 & \text{if } i \in F,\ i \neq f_j \\ -b_{k,f_j} & \text{if } i \in D,\ i = d_k \end{cases}.$$

Then the set of solutions to the system of equations represented by the vector equation

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{c} + x_{f_1}\mathbf{u}_1 + x_{f_2}\mathbf{u}_2 + x_{f_3}\mathbf{u}_3 + \cdots + x_{f_{n-r}}\mathbf{u}_{n-r}$$

is equal to the set of solutions of $\mathcal{LS}(A,\, \mathbf{b})$. □

Proof We are being asked to prove that two systems of equations are equivalent, that is, they have identical solution sets. 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 [37]). We will now show that the equations in the conclusion of the proof are either always true, or are simple rearrangements of the equations in the system with B as its augmented matrix. This will then establish that all three systems are equivalent to each other.

Suppose that $i \in F$, so $i = f_k$ for a particular choice of k, $1 \leq k \leq n-r$. Consider the equation given by entry i of the vector equality.

$$\begin{aligned} x_i &= c_i + x_{f_1}u_{i1} + x_{f_2}u_{i2} + x_{f_3}u_{i3} + \cdots + x_{f_{n-r}}u_{i,n-r} \\ x_i &= c_{f_k} + x_{f_1}u_{f_k1} + x_{f_2}u_{f_k2} + x_{f_3}u_{f_k3} + \cdots + x_{f_k}u_{f_kf_k} + \cdots + x_{f_{n-r}}u_{f_k,n-r} \\ x_i &= 0 + x_{f_1}(0) + x_{f_2}(0) + x_{f_3}(0) + \cdots + x_{f_k}(1) + \cdots + x_{f_{n-r}}(0) = x_{f_k} \\ x_i &= x_i. \end{aligned}$$

This means that equality of the two vectors in entry i represents the equation $x_i = x_i$ when $i \in F$. Since this equation is always true, it does not restrict the possibilities for the solution set.

Now consider the i-th entry, when $i \in D$, and suppose that $i = d_k$, for some particular choice of k, $1 \leq k \leq r$. Consider the equation given by entry i of the vector equality.

$$\begin{aligned} x_i &= c_i + x_{f_1}u_{i1} + x_{f_2}u_{i2} + x_{f_3}u_{i3} + \cdots + x_{f_{n-r}}u_{i,n-r} \\ x_i &= c_{d_k} + x_{f_1}u_{d_k1} + x_{f_2}u_{d_k2} + x_{f_3}u_{d_k3} + \cdots + x_{f_{n-r}}u_{d_k,n-r} \\ x_i &= b_{k,n+1} + x_{f_1}(-b_{k,f_1}) + x_{f_2}(-b_{k,f_2}) + x_{f_3}(-b_{k,f_3}) + \cdots + x_{f_{n-r}}(-b_{k,f_{n-r}}) \\ x_i &= b_{k,n+1} - \left(b_{k,f_1}x_{f_1} + b_{k,f_2}x_{f_2} + b_{k,f_3}x_{f_3} + \cdots + b_{k,f_{n-r}}x_{f_{n-r}}\right) \end{aligned}$$

Rearranging, this becomes,

$$x_i + b_{k,f_1}x_{f_1} + b_{k,f_2}x_{f_2} + b_{k,f_3}x_{f_3} + \cdots + b_{k,f_{n-r}}x_{f_{n-r}} = b_{k,n+1}.$$

This is exactly the equation represented by row k of the matrix B. So the equations represented by the vector equality in the conclusion are exactly the equations represented by the matrix B, along with additional equations of the form $x_i = x_i$ that are always true. So the solution sets will be identical. ■

Theorem VFSLS [104] formalizes what happened in the three steps of Example VFSAD [103]. 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 [103] when I need to describe an infinite solution set. So lets practice some more, but with a bigger example.

Example VFSAI
Vector form of solutions for Archetype I
Archetype I [510] is a linear system of $m = 4$ equations in $n = 7$ variables. Row-reducing the augmented matrix yields

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

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.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix} + x_2 \begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix} + x_5 \begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix} + x_6 \begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix} + x_7 \begin{bmatrix} \\ \\ \\ \\ \\ \\ \end{bmatrix}$$

Step 2. For each free variable, use 0's and 1's to ensure equality for the corresponding entry of the 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.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} \\ 0 \\ \\ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} \\ 1 \\ \\ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} \\ 0 \\ \\ \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} \\ 0 \\ \\ \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} \\ 0 \\ \\ \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

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.

$$x_1 = 4 - 4x_2 - 2x_5 - 1x_6 + 3x_7 \quad \Rightarrow$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ \\ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -4 \\ 1 \\ \\ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -2 \\ 0 \\ \\ \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} -1 \\ 0 \\ \\ \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 3 \\ 0 \\ \\ \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$x_3 = 2 + 0x_2 - x_5 + 3x_6 - 5x_7 \quad \Rightarrow$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \\ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -4 \\ 1 \\ 0 \\ \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -2 \\ 0 \\ -1 \\ \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} -1 \\ 0 \\ 3 \\ \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 3 \\ 0 \\ -5 \\ \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$x_4 = 1 + 0x_2 - 2x_5 + 6x_6 - 6x_7 \quad \Rightarrow$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -2 \\ 0 \\ -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} -1 \\ 0 \\ 3 \\ 6 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_7 \begin{bmatrix} 3 \\ 0 \\ -5 \\ -6 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

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 [510]. (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 [510] you know that's your cue to go work through Archetype J [515] 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 [525] 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 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

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

and we see $r = 3$ nonzero rows. The columns with leading 1's are $D = \{1, 2, 3\}$ 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.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix} + x_4 \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix} + x_5 \begin{bmatrix} \\ \\ \\ \\ \\ \end{bmatrix}$$

Step 2. For each free variable, use 0's and 1's to ensure equality for the corresponding entry of the 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.

$$x_1 = 0 - 1x_4 + 2x_5 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ \\ \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ \\ \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ \\ \\ 0 \\ 1 \end{bmatrix}$$

$$x_2 = 0 + 2x_4 - 2x_5 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 2 \\ \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ -2 \\ \\ 0 \\ 1 \end{bmatrix}$$

$$x_3 = 0 - 2x_4 + 1x_5 \quad \Rightarrow \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 2 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 2 \\ -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

The vector $\mathbf{c}$ 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 $\mathbf{c}$ 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 $\mathbf{u}_1 = \begin{bmatrix}-1\\2\\-2\\1\\0\end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix}2\\-2\\1\\0\\1\end{bmatrix}$, 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. ⊚

Subsection URREF Uniqueness of Reduced Row-Echelon Form

We are now in a position to establish that the reduced row-echelon form of a matrix is unique. Going forward, we will emphasize the point-of-view that a matrix is a collection of columns. But there are two occasions when we need to work carefully with the rows of a matrix. This is the first such occasion. We could define something called a **row vector** that would equal a given row of a matrix, and might be written as a horizontal list. Then we could define vector equality, the basic operations of vector addition and scalar multiplication, followed by a definition of a linear combination of row vectors. We will not incur the overhead of stating all these definitions, but will instead convert the rows of a matrix to column vectors and use our definitions that are already in place. This was our reason for delaying this proof until now. Remind yourself as you work through this proof that it only relies only on the definition of equivalent matrices, reduced row-echelon form and linear combinations. So in particular, we are not guilty of circular reasoning. Should we have defined vector operations and linear combinations just prior to discussing reduced row-echelon form, then the following proof of uniqueness could have been presented at that time. OK, here we go.

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 Denote the pivot columns of B as $D = \{d_1, d_2, d_3, \ldots, d_r\}$ and the pivot columns of C as $D' = \{d'_1, d'_2, d'_3, \ldots, d'_{r'}\}$ (Notation RREFA [51]). We begin by showing that $D = D'$.

For both B and C, we can take the elements of a row of the matrix and use them to construct a column vector. We will denote these by $\mathbf{b}_i$ and $\mathbf{c}_i$, respectively, $1 \leq i \leq m$. Since B and C are both row-equivalent to A, there is a sequence of row operations that will convert B to C, and vice-versa, since row operations are reversible. If we can convert B into C via a sequence of row operations, then any row of C expressed as a column vector, say $\mathbf{c}_k$, is a linear combination of the column vectors derived from the rows of B, $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \ldots, \mathbf{b}_m\}$. Similarly, any row of B is a linear combination of the set of rows of C. Our principal device in this proof is to carefully analyze individual entries of vector equalities between a single row of either B or C and a linear combination of the rows of the other matrix.

Lets first show that $d_1 = d'_1$. Suppose that $d_1 < d'_1$. We can write the first row of B as a linear combination of the rows of C, that is, there are scalars $a_1, a_2, a_3, \ldots, a_m$ such that

$$\mathbf{b}_1 = a_1\mathbf{c}_1 + a_2\mathbf{c}_2 + a_3\mathbf{c}_3 + \cdots + a_m\mathbf{c}_m$$

Consider the entry in location d_1 on both sides of this equality. Since B is in reduced row-echelon form (Definition RREF [38]) we find a one in $\mathbf{b}_1$ on the left. Since $d_1 < d'_1$, and C is in reduced row-echelon form (Definition RREF [38]) each vector $\mathbf{c}_i$ has a zero in location d_1, and therefore the linear combination on the right also has a zero in location d_1. This is a contradiction, so we know that $d_1 \geq d'_1$. By an entirely similar argument, we could conclude that $d_1 \leq d'_1$. This means that $d_1 = d'_1$.

Suppose that we have determined that $d_1 = d'_1$, $d_2 = d'_2$, $d_3 = d'_3$, $\ldots$, $d_k = d'_k$. Lets now show that $d_{k+1} = d'_{k+1}$. To achieve a contradiction, suppose that $d_{k+1} < d'_{k+1}$. Row $k+1$ of B is a linear combination of the rows of C, so there are scalars $a_1, a_2, a_3, \ldots, a_m$ such that

$$\mathbf{b}_{k+1} = a_1\mathbf{c}_1 + a_2\mathbf{c}_2 + a_3\mathbf{c}_3 + \cdots + a_m\mathbf{c}_m$$

Since B is in reduced row-echelon form (Definition RREF [38]), the entries of $\mathbf{b}_{k+1}$ in locations $d_1, d_2, d_3, \ldots, d_k$ are all zero. Since C is in reduced row-echelon form (Definition RREF [38]), location d_i of $\mathbf{c}_i$ is one for each $1 \leq i \leq k$. The equality of these vectors in locations $d_1, d_2, d_3, \ldots, d_k$ then implies that $a_1 = 0$, $a_2 = 0$, $a_3 = 0$, $\ldots$, $a_k = 0$.

Now consider location d_{k+1} in this vector equality. The vector $\mathbf{b}_{k+1}$ on the left is one in this location since B is in reduced row-echelon form (Definition RREF [38]). Vectors $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \ldots, \mathbf{c}_k$, are multiplied by zero scalars in the linear combination on the right. The remaining vectors, $\mathbf{c}_{k+1}, \mathbf{c}_{k+2}, \mathbf{c}_{k+3}, \ldots, \mathbf{c}_m$ each has a zero in location d_{k+1} since $d_{k+1} < d'_{k+1}$ and C is in reduced row-echelon form (Definition RREF [38]). So the right hand side of the vector equality is zero in location d_{k+1}, a contradiction. Thus $d_{k+1} \geq d'_{k+1}$. By an entirely similar argument, we could conclude that $d_{k+1} \leq d'_{k+1}$, and therefore $d_{k+1} = d'_{k+1}$.

Now we establish that $r = r'$. Suppose that $r < r'$. By the arguments above we can show that $d_1 = d'_1$, $d_2 = d'_2$, $d_3 = d'_3$, $\ldots$, $d_r = d'_r$. Row r' of C is a linear combination of the r non-zero rows of B, so there are scalars $a_1, a_2, a_3, \ldots, a_r$ so that

$$\mathbf{c}_{r'} = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + a_3\mathbf{b}_3 + \cdots + a_r\mathbf{b}_r$$

Locations $d_1, d_2, d_3, \ldots, d_r$ of $\mathbf{c}_{r'}$ are all zero since $r < r'$ and C is in reduced row-echelon form (Definition RREF [38]). For a given index i, $1 \leq i \leq r$, the vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \ldots, \mathbf{b}_r$ have zeros in location d_i, except that the vector $\mathbf{b}_i$ is one in location d_i since B is in reduced row-echelon form (Definition RREF [38]). This consideration of location d_i implies that $a_i = 0$, $1 \leq i \leq r$. With all the scalars in the linear combination equal to zero, we conclude that $\mathbf{c}_{r'} = \mathbf{0}$, contradicting the existence of a leading 1 in $\mathbf{c}_{r'}$. So $r \geq r'$. By a similar argument, we conclude that $r \leq r'$ and therefore $r = r'$. Thus $D = D'$.

To finally show that $B = C$, we will show that the rows of the two matrices are equal. Row k of C, $\mathbf{c}_k$, is a linear combination of the r non-zero rows of B, so there are scalars $a_1, a_2, a_3, \ldots, a_r$ such that

$$\mathbf{c}_k = a_1\mathbf{b}_1 + a_2\mathbf{b}_2 + a_3\mathbf{b}_3 + \cdots + a_r\mathbf{b}_r$$

Because C is in reduced row-echelon form (Definition RREF [38]), location d_i of $\mathbf{c}_k$ is zero for $1 \leq i \leq r$, except in location d_k where the entry is one. In the linear combination on the right of the vector equality, the vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \ldots, \mathbf{b}_r$ have zeros in location d_i, except that $\mathbf{b}k$ has a one in location d_i, since B is in reduced row-echelon form (Definition RREF [38]). This implies that $a_1 = 0$, $a_2 = 0$, $\ldots$, $a_{k-1} = 0$, $a_{k+1} = 0$, $a_{k+2} = 0$, $\ldots$, $a_r = 0$ and $a_k = 1$. Then the vector equality reduces to simply $\mathbf{c}_k = \mathbf{b}_k$. Since k was arbitrary, B and C have equal rows and so are equal matrices. ■

Subsection READ Reading Questions

1. Earlier, a reading question asked you to solve the system of equations

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 0 \\ x_1 + 2x_2 + x_3 &= 3 \\ x_1 + 3x_2 + 3x_3 &= 7 \end{aligned}$$

   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.

$$\begin{bmatrix} \boxed{1} & 3 & 0 & 6 & 0 & 9 \\ 0 & 0 & \boxed{1} & -2 & 0 & -8 \\ 0 & 0 & 0 & 0 & \boxed{1} & 3 \end{bmatrix}$$

Subsection EXC
Exercises

C40 Contributed by Robert Beezer
Find the vector form of the solutions to the system of equations below.

$$\begin{aligned}
2x_1 - 4x_2 + 3x_3 + x_5 &= 6 \\
x_1 - 2x_2 - 2x_3 + 14x_4 - 4x_5 &= 15 \\
x_1 - 2x_2 + x_3 + 2x_4 + x_5 &= -1 \\
-2x_1 + 4x_2 - 12x_4 + x_5 &= -7
\end{aligned}$$

Solution [115]

Subsection SOL Solutions

C40 Exercise [113] Contributed by Robert Beezer
Row-reduce the augmented matrix representing this system, to find

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

The system is consistent (no leading one in column 6, Theorem RCLS [56]). x_2 and x_4 are the free variables. Now apply Theorem VFSLS [104], or follow the three-step process of Example VFSAD [103], Example VFSAI [106], or Example VFSAL [108] to obtain

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \\ -5 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -6 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix}$$

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.

Subsection SSV
Span of a Set of Vectors

In Example VFSAL [108] 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 $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_t\}$, their **span**, $\mathcal{S}p(S)$, is the set of all possible linear combinations of $\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_t$. Symbolically,

$$\begin{aligned}\mathcal{S}p(S) &= \{\,\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_t\mathbf{u}_t \mid \alpha_i \in \mathbb{C},\, 1 \le i \le t\}\\ &= \left\{\sum_{i=1}^{t}\alpha_i\mathbf{u}_i \,\middle|\, \alpha_i \in \mathbb{C},\, 1 \le i \le t\right\}\end{aligned} \qquad \triangle$$

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 (t of them to be precise), and use this finite set to describe an infinite set of vectors. We will see this construction repeatedly, so lets 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 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 $U = \mathcal{S}p(S)$. 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 [473]. 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 $\mathcal{S}p(S)$, since it is a linear combination of $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$. We write this succinctly as $\mathbf{v} \in \mathcal{S}p(S)$. 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 $\mathcal{S}p(S)$. A slightly different question arises when you are handed a vector of the correct size and asked if it is an element of $\mathcal{S}p(S)$. For example, is $\mathbf{w} = \begin{bmatrix}1\\8\\5\end{bmatrix}$ in $\mathcal{S}p(S)$? More succinctly, $\mathbf{w} \in \mathcal{S}p(S)$?

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 [101] solutions to this vector equality are solutions to the system of equations

$$\begin{aligned} \alpha_1 - \alpha_2 + 2\alpha_3 &= 1\\ 2\alpha_1 + \alpha_2 + \alpha_3 &= 8\\ \alpha_1 + \alpha_2 &= 5. \end{aligned}$$

Building the augmented matrix for this linear system, and row-reducing, gives

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

This system has infinitely many solutions (there's a free variable), but all we need is one. The solution,

$$\alpha_1 = 2 \qquad \alpha_2 = 3 \qquad \alpha_3 = 1$$

tells us that

$$(2)\mathbf{u}_1 + (3)\mathbf{u}_2 + (1)\mathbf{u}_3 = \mathbf{w}$$

so we are convinced that $\mathbf{w}$ really is in $\mathcal{S}p(S)$. Lets ask the same question again, but this time with $\mathbf{y} = \begin{bmatrix}2\\4\\3\end{bmatrix}$, i.e. is $\mathbf{y} \in \mathcal{S}p(S)$?

So we'll look for scalars α_1, α_2, α_3 so that

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 = \mathbf{y}.$$

By Theorem SLSLC [101] this linear combination becomes the system of equations

$$\begin{aligned} \alpha_1 - \alpha_2 + 2\alpha_3 &= 2 \\ 2\alpha_1 + \alpha_2 + \alpha_3 &= 4 \\ \alpha_1 + \alpha_2 &= 3. \end{aligned}$$

Building the augmented matrix for this linear system, and row-reducing, gives

$$\begin{bmatrix} \boxed{1} & 0 & 1 & 0 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

This system is inconsistent (there's a leading 1 in the last column), 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 $\mathbf{y}$. More precisely, $\mathbf{y} \notin \mathcal{S}p(S)$.

There are three things to observe in this example. (1) It is easy to construct vectors in $\mathcal{S}p(S)$. (2) It is possible that some vectors are in $\mathcal{S}p(S)$ (e.g. $\mathbf{w}$), while others are not (e.g. $\mathbf{y}$). (3) Deciding if a given vector is in $\mathcal{S}p(S)$ 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 [473]. Study the determination that $\mathbf{v} \in \mathcal{S}p(S)$ and see if you can connect it with some of the other properties of Archetype A [473]. ⊚

Lets do a similar example to Example SCAA [117], only now with Archetype B [478].

Example SCAB
Span of the columns of Archetype B
Begin with the finite set of three vectors of size 3

$$R = \{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3\} = \left\{ \begin{bmatrix} -7 \\ 5 \\ 1 \end{bmatrix},\, \begin{bmatrix} -6 \\ 5 \\ 0 \end{bmatrix},\, \begin{bmatrix} -12 \\ 7 \\ 4 \end{bmatrix} \right\}$$

and consider the infinite set $V = \mathcal{S}p(R)$. The vectors of R have been chosen as the three columns of the coefficient matrix in Archetype B [478]. 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 $\mathcal{S}p(R)$, since it is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$. In other words, $\mathbf{x} \in \mathcal{S}p(R)$. Try some different values of α_1, α_2, α_3 yourself, and see what vectors you can create as elements of $\mathcal{S}p(R)$.

Now ask if a given vector is an element of $\mathcal{S}p(R)$. For example, is $\mathbf{z} = \begin{bmatrix} -33 \\ 24 \\ 5 \end{bmatrix}$ in $\mathcal{S}p(R)$? Is $\mathbf{z} \in \mathcal{S}p(R)$?

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 [101] this linear combination becomes the system of equations

$$\begin{aligned} -7\alpha_1 - 6\alpha_2 - 12\alpha_3 &= -33 \\ 5\alpha_1 + 5\alpha_2 + 7\alpha_3 &= 24 \\ \alpha_1 + 4\alpha_3 &= 5. \end{aligned}$$

Building the augmented matrix for this linear system, and row-reducing, gives

$$\begin{bmatrix} \boxed{1} & 0 & 0 & -3 \\ 0 & \boxed{1} & 0 & 5 \\ 0 & 0 & \boxed{1} & 2 \end{bmatrix}.$$

This system has a unique solution,

$$\alpha_1 = -3 \qquad \alpha_2 = 5 \qquad \alpha_3 = 2$$

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 $\mathcal{S}p(R)$.

There is no point in replacing $\mathbf{z}$ with another vector and doing this again. A question about membership in $\mathcal{S}p(R)$ 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 NSMUS [79], it is guaranteed to have a solution. (This solution is unique, but that's not important 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 $\mathcal{S}p(R)$. Stated differently, every vector of size 3 is in $\mathcal{S}p(R)$, or $\mathcal{S}p(R) = \mathbb{C}^3$.

Compare this example with Example SCAA [117], and see if you can connect $\mathbf{z}$ with some aspects of the write-up for Archetype B [478]. ⊚

Subsection SSNS
Spanning Sets of Null Spaces

We saw in Example VFSAL [108] that when a system of equations is homogeneous the solution set can be expressed in the form described by Theorem VFSLS [104] where the vector $\mathbf{c}$ 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 $D = \{d_1,\,d_2,\,d_3,\,\ldots,\,d_r\}$ 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 $\mathbf{u}_j = (u_{ij})$, $1 \leq j \leq n-r$ of size n as

$$u_{ij} = \begin{cases} 1 & \text{if } i \in F,\, i = f_j \\ 0 & \text{if } i \in F,\, i \neq f_j \\ -b_{k,f_j} & \text{if } i \in D,\, i = d_k \end{cases}.$$

Then the null space of A is given by

$$\mathcal{N}(A) = \mathcal{S}p(\{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_{n-r}\})\,. \qquad \square$$

Proof Consider the homogeneous system with A as a coefficient matrix, $\mathcal{LS}(A,\,\mathbf{0})$. Its set of solutions is, by definition, the null space of A, $N(A)$. Row-reducing the augmented matrix of this homogeneous system will create the row-equivalent matrix B'. Row-reducing the augmented matrix that has a final column of all zeros, yields B', which is the matrix B, along with an additional column (index $n+1$) that is still totally zero.

Now apply Theorem VFSLS [104], noting that our homogeneous system is consistent (Theorem HSC [68]). The vector $\mathbf{c}$ 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' these entries in the final column are all zero. So $\mathbf{c} = \mathbf{0}$. This says that a solution of the homogeneous system is of the form

$$\mathbf{x} = \mathbf{c} + x_{f_1}\mathbf{u}_1 + x_{f_2}\mathbf{u}_2 + x_{f_3}\mathbf{u}_3 + \cdots + x_{f_{n-r}}\mathbf{u}_{n-r} = x_{f_1}\mathbf{u}_1 + x_{f_2}\mathbf{u}_2 + x_{f_3}\mathbf{u}_3 + \cdots + x_{f_{n-r}}\mathbf{u}_{n-r}$$

where the free variables x_{f_j} can each take on any value. Rephrased this says

$$\begin{aligned} \mathcal{N}(A) &= \left\{ x_{f_1}\mathbf{u}_1 + x_{f_2}\mathbf{u}_2 + x_{f_3}\mathbf{u}_3 + \cdots + x_{f_{n-r}}\mathbf{u}_{n-r} \,\middle|\, x_{f_1},\,x_{f_2},\,x_{f_3},\,\ldots,\,x_{f_{n-r}} \in \mathbb{C} \right\} \\ &= \mathcal{S}p(\{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_{n-r}\})\,. \end{aligned} \qquad \blacksquare$$

Here's an example that will exercise the span construction and Theorem SSNS [121], 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 $W = \mathcal{S}p(T)$. The vectors of T have been chosen as the four columns of the coefficient matrix in Archetype D [487]. Check that the vector

$$\mathbf{u}_2 = \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

is a solution to the homogeneous system $\mathcal{LS}(D, \mathbf{0})$ (it is the second vector of the spanning set for the null space of the coefficient matrix D, as described in Theorem SSNS [121]). Applying Theorem SLSLC [101], 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,

$$\begin{aligned} 5\mathbf{w}_1 + (-4)\mathbf{w}_2 + 6\mathbf{w}_3 + (-3)\mathbf{w}_4 &= 5\mathbf{w}_1 + (-4)\mathbf{w}_2 + 6\mathbf{w}_3 + (-3)\left((-2)\mathbf{w}_1 + (-3)\mathbf{w}_2\right) \\ &= 5\mathbf{w}_1 + (-4)\mathbf{w}_2 + 6\mathbf{w}_3 + (6\mathbf{w}_1 + 9\mathbf{w}_2) \\ &= 11\mathbf{w}_1 + 5\mathbf{w}_2 + 6\mathbf{w}_3. \end{aligned}$$

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 (Technique SE [21]) would now allow us to conclude that this reduction is possible for any vector in W, so

$$W = \mathcal{S}p(\{\mathbf{w}_1,\, \mathbf{w}_2,\, \mathbf{w}_3\}).$$

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

$$\mathbf{u}_1 = \begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

is a solution to the homogeneous system $\mathcal{LS}(D, \mathbf{0})$ (it is the first vector of the spanning set for the null space of the coefficient matrix D, as described in Theorem SSNS [121]). Applying Theorem SLSLC [101], 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 would show that

$$W = \mathcal{S}p(\{\mathbf{w}_1,\, \mathbf{w}_2\})\,.$$

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*. ⊚

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 $W = \mathcal{S}p(S)$ 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 that spans the null space of A, $\mathcal{N}(A)$. That is, S should be such that $\mathcal{S}p(S) = \mathcal{N}(A)$. (See Theorem SSNS [121].)

$$A = \begin{bmatrix} 1 & 3 & 1 & 9 \\ 2 & 1 & -3 & 8 \\ 1 & 1 & -1 & 5 \end{bmatrix}$$

Subsection EXC
Exercises

C40 Contributed by Robert Beezer

Suppose that $S = \left\{ \begin{bmatrix} 2 \\ -1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ -2 \\ 1 \end{bmatrix} \right\}$. Let $W = \mathcal{S}p(S)$ 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. Solution [127]

C41 Contributed by Robert Beezer

Suppose that $S = \left\{ \begin{bmatrix} 2 \\ -1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ -2 \\ 1 \end{bmatrix} \right\}$. Let $W = \mathcal{S}p(S)$ 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. Solution [127]

C60 Contributed by Robert Beezer

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, $\mathcal{S}p(S) = \mathcal{N}(A)$.

$$A = \begin{bmatrix} 1 & 1 & 6 & -8 \\ 1 & -2 & 0 & 1 \\ -2 & 1 & -6 & 7 \end{bmatrix}$$

Solution [128]

Subsection SOL
Solutions

C40 Exercise [125] Contributed by Robert Beezer
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 [101] 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

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

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 Exercise [125] Contributed by Robert Beezer
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 [101] 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

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \end{bmatrix}$$

With a leading 1 in the last column of this matrix (Theorem RCLS [56]) 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 $\mathbf{y}$ is in W. So $\mathbf{y} \notin W$.

C60 Exercise [125] Contributed by Robert Beezer
Theorem BNS [138] 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,

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

Moving to the vector form of the solutions (Theorem VFSLS [104]), with free variables x_3 and x_4, solutions to the consistent system (it is homogeneous, Theorem HSC [68]) 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 [138] guarantees the set has the desired properties.

Section LI
Linear Independence

Subsection LIV
Linearly Independent Vectors

Theorem SLSLC [101] 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 [121] 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 $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_n\}$, an equation of the form

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n = \mathbf{0}$$

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. △

Definition LICV
Linear Independence of Column Vectors
The set of vectors $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_n\}$ 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. △

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 [101] 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,

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

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 [101], 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 [101] tells us that we can find such solutions as solution to the homogeneous system $\mathcal{LS}(B,\,\mathbf{0})$ where the coefficient matrix has these four vectors as columns,

$$B=\begin{bmatrix}2&1&2&-6\\-1&2&1&7\\3&-1&-3&-1\\1&5&6&1\\2&2&1&1\end{bmatrix}.$$

Row-reducing this coefficient matrix yields,

$$\begin{bmatrix}\boxed{1}&0&0&0\\0&\boxed{1}&0&0\\0&0&\boxed{1}&0\\0&0&0&\boxed{1}\\0&0&0&0\end{bmatrix}.$$

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 [129] and Example LIS [131] 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 $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n\}$ 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 (⇐) 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 [101], this means that the only relation of linear dependence on S is the trivial one. So S is linearly independent.

(⇒) 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 [68]), and so must have infinitely many solutions (Theorem PSSLS [58]). One of these infinitely many solutions must be nontrivial (in fact, almost all of them are), so choose one. By Theorem SLSLC [101] 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 [132] 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 another example.

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 [132], 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 [69] tells us that since the system is homogeneous with $n = 9$ variables in $m = 4$ equations, and $n > m$, there must be infinitely many solutions. Since there is not a unique solution, Theorem LIVHS [132] says the set is linearly dependent. ⊚

The situation in Example LLDS [132] is slick enough to warrant formulating as a theorem.

Theorem MVSLD
More Vectors than Size implies Linear Dependence
Suppose that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_n\}$ 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 [69] this system has infinitely many solutions. Since the system does not have a unique solution, Theorem LIVHS [132] says the columns of A form a linearly dependent set, which is the desired conclusion. ■

As an equivalence, Theorem LIVHS [132] gives us a straightforward way to determine if a set of vectors is linearly independent or dependent. We can improve on it just slightly, so we will state this corollary as a theorem.

Theorem LIVRN
Linearly Independent Vectors, r and n
Suppose that A is an $m \times n$ matrix and $S = \{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n\}$ 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 [132] 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 [68]) we can apply Theorem CSRN [57] to see that the solution is unique exactly when $n = r$. ■

Subsection LDSS
Linearly Dependent Sets 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 [134]. 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).

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 [134]. 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 to go about it — 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 $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_n\}$ is a set of vectors. Then S is a linearly dependent set if and only if there is an index t, $1 \leq t \leq n$ such that $\mathbf{u_t}$ is a linear combination of the vectors $\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_{t-1},\, \mathbf{u}_{t+1},\, \ldots,\, \mathbf{u}_n$. □

Proof ($\Rightarrow$) Suppose that S is linearly dependent, so there is a nontrivial relation of linear dependence,

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n = \mathbf{0}.$$

Since the α_i cannot all be zero, choose one, say α_t, that is nonzero. Then,

$$-\alpha_t\mathbf{u}_t = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_{t-1}\mathbf{u}_{t-1} + \alpha_{t+1}\mathbf{u}_{t+1} + \cdots + \alpha_n\mathbf{u}_n$$

and we can multiply by $\frac{-1}{\alpha_t}$ since $\alpha_t \neq 0$,

$$\mathbf{u}_t = \frac{-\alpha_1}{\alpha_t}\mathbf{u}_1 + \frac{-\alpha_2}{\alpha_t}\mathbf{u}_2 + \frac{-\alpha_3}{\alpha_t}\mathbf{u}_3 + \cdots + \frac{-\alpha_{t-1}}{\alpha_t}\mathbf{u}_{t-1} + \frac{-\alpha_{t+1}}{\alpha_t}\mathbf{u}_{t+1} + \cdots + \frac{-\alpha_n}{\alpha_t}\mathbf{u}_n.$$

Since the values of $\frac{\alpha_i}{\alpha_t}$ are again scalars, we have expressed $\mathbf{u}_t$ as the desired linear combination.

($\Leftarrow$) Suppose that the vector $\mathbf{u}_t$ is a linear combination of the other vectors in S. Write this linear combination as

$$\beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \beta_3\mathbf{u}_3 + \cdots + \beta_{t-1}\mathbf{u}_{t-1} + \beta_{t+1}\mathbf{u}_{t+1} + \cdots + \beta_n\mathbf{u}_n = \mathbf{u_t}$$

and move $\mathbf{u}_t$ to the other side of the equality

$$\beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \beta_3\mathbf{u}_3 + \cdots + \beta_{t-1}\mathbf{u}_{t-1} + (-1)\mathbf{u}_t + \beta_{t+1}\mathbf{u}_{t+1} + \cdots + \beta_n\mathbf{u}_n = \mathbf{0}.$$

Then the scalars $\beta_1,\, \beta_2,\, \beta_3,\, \ldots,\, \beta_t = -1,\, \ldots,\, \beta_n$ provide a *nontrivial* linear combination of the vectors in S, thus establishing that S is a linearly dependent set. ■

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 [121], 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 $V = \mathcal{S}p(R)$.

To employ Theorem LIVHS [132], 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 $\mathcal{LS}(D,\, \mathbf{0})$,

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 4 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We can find infinitely many solutions to this system, most of them nontrivial, and we choose anyone we like to build a relation of linear dependence on R. Lets begin with $x_4 = 1$, to find the solution

$$\begin{bmatrix} -4 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

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 [134] 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 $\mathbf{v}_2$, lets 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

$$V = \mathcal{S}p(R) = \mathcal{S}p(\{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3,\, \mathbf{v}_4\}) = \mathcal{S}p(\{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_4\})$$

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 Technique SE [21] to establish it carefully. Let $R' = \{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_4\}$ and $V' = \mathcal{S}p(R')$. 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 $\mathbf{v}$ from V. Then there are scalars α_1, α_2, α_3, α_4 so that

$$\begin{aligned}\mathbf{v} &= \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 + \alpha_4\mathbf{v}_4\\ &= \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\left((-4)\mathbf{v}_1 + 1\mathbf{v}_4\right) + \alpha_4\mathbf{v}_4\\ &= \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \left((-4\alpha_3)\mathbf{v}_1 + \alpha_3\mathbf{v}_4\right) + \alpha_4\mathbf{v}_4\\ &= \left(\alpha_1 - 4\alpha_3\right)\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \left(\alpha_3 + \alpha_4\right)\mathbf{v}_4.\end{aligned}$$

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 (its 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 LINSM
Linear Independence and NonSingular Matrices

We will now specialize to sets of n vectors from $\mathbb{C}^n$. This will put Theorem MVSLD [133] off-limits, while Theorem LIVHS [132] will involve square matrices. Lets begin by contrasting Archetype A [473] and Archetype B [478].

Example LDCAA
Linearly dependent columns in Archetype A
Archetype A [473] 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 [76] we know that A is singular. According to the definition of nonsingular matrices, Definition NM [75], the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has infinitely many solutions. So by Theorem LIVHS [132], the columns of A form a linearly dependent set. ⊚

Example LICAB
Linearly independent columns in Archetype B

Archetype B [478] 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 NS [76] we know that B is nonsingular. According to the definition of nonsingular matrices, Definition NM [75], the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution. So by Theorem LIVHS [132], the columns of B form a linearly independent set. ⊚

That Archetype A [473] and Archetype B [478] 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 NSME1 [82].

Theorem NSLIC
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. □

Proof This is a proof where we can chain together equivalences, rather than proving the two halves separately. By definition, A is nonsingular if and only if the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution. Theorem LIVHS [132] then says that the system $\mathcal{LS}(A, \mathbf{0})$ has a unique solution if and only if the columns of A are a linearly independent set. ■

Here's an update to Theorem NSME1 [82].

Theorem NSME2
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 NSLIC [137] is yet another equivalence for a nonsingular matrix, so we can add it to the list in Theorem NSME1 [82]. ■

Subsection NSSLI
Null Spaces, Spans, Linear Independence

In Subsection SS.SSNS [120] we proved Theorem SSNS [121] 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 seen in Theorem DLDS [134] and Example RSC5 [134], 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 [121] 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 $\mathbf{u}_j$, $1 \leq j \leq n-r$ first appear in the vector form of solutions to arbitrary linear systems (Theorem VFSLS [104]). The exact same vectors appear again in the span construction in the conclusion of Theorem SSNS [121]. Since this second theorem specializes to homogeneous systems the only real difference is that the vector $\mathbf{c}$ in Theorem VFSLS [104] is the zero vector for a homogeneous system. Finally, Theorem BNS [138] will now show that these same vectors are a linearly independent set.

The proof is really quite straightforward, and relies on the “pattern” of zeros and ones that arise in the vectors $\mathbf{u}_i$, $1 \leq i \leq n-r$ in the entries that correspond to the free variables. So take a look at Example VFSAD [103], Example VFSAI [106] and Example VFSAL [108], especially during the conclusion of Step 2 (temporarily ignore the construction of the constant vector, $\mathbf{c}$). It is a good exercise in showing 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 $\mathbf{z}_j = (z_{ij})$, $1 \leq j \leq n-r$ of size n as

$$z_{ij} = \begin{cases} 1 & \text{if } i \in F,\ i = f_j \\ 0 & \text{if } i \in F,\ i \neq f_j \\ -b_{k,f_j} & \text{if } i \in D,\ i = d_k \end{cases}.$$

Define the set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_{n-r}\}$. Then

1. $\mathcal{N}(A) = \mathcal{S}p(S)$.
2. S is a linearly independent set. □

Proof Notice first that the vectors $\mathbf{z}_j = (z_{ij})$, $1 \leq j \leq n-r$ are defined in exactly the same way that the vectors $\mathbf{u}_j = (u_{ij})$, $1 \leq j \leq n-r$ of Theorem SSNS [121] are defined. Other than this cosmetic change in the names of these vectors, the hypotheses of Theorem SSNS [121] are the same as the hypotheses of the theorem we are currently proving. So it is then simply the conclusion of Theorem SSNS [121] that tells us that $\mathcal{N}(A) = \mathcal{S}p(S)$. That was the easy half, but the second part is not much harder.

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

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_{n-r}\mathbf{u}_{n-r} = \mathbf{0}.$$

For each j, $1 \leq j \leq n-r$, consider the entry of the vectors on both sides of this equality in position f_j. On the right, it is easy since the zero vector has a zero in each entry. On the left we find,

$$\begin{aligned}
&\alpha_1\mathbf{z}_{f_j1} + \alpha_2\mathbf{z}_{f_j2} + \alpha_3\mathbf{z}_{f_j3} + \cdots + \alpha_j\mathbf{z}_{f_j,j} + \cdots + \alpha_{n-r}\mathbf{z}_{f_j,n-r} \\
&= \alpha_1(0) + \alpha_2(0) + \alpha_3(0) + \cdots + \alpha_j(1) + \cdots + \alpha_{n-r}(0) \\
&= \alpha_j
\end{aligned}$$

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 the set is linearly independent, as desired. ■

Example NSLIL
Null space spanned by linearly independent set, Archetype L
In Example VFSAL [108] we previewed Theorem SSNS [121] by finding a set of two vectors such that their span was the null space for the matrix in Archetype L [525]. Writing the matrix as L, we have

$$N(L) = \mathcal{S}p\left(\left\{\begin{bmatrix}-1\\2\\-2\\1\\0\end{bmatrix},\ \begin{bmatrix}2\\-2\\1\\0\\1\end{bmatrix}\right\}\right).$$

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?

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?

3. Based on your answer to the previous question, is the matrix below singular or nonsingular?

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

Subsection EXC
Exercises

Determine if the sets of vectors in Exercises C20–C22 are linearly independent or linearly dependent. **C20** Contributed by Robert Beezer

$$\left\{\begin{bmatrix}1\\-2\\1\end{bmatrix},\,\begin{bmatrix}2\\-1\\3\end{bmatrix},\,\begin{bmatrix}1\\5\\0\end{bmatrix}\right\}$$

Solution [143]

C21 Contributed by Robert Beezer

$$\left\{\begin{bmatrix}-1\\2\\4\\2\end{bmatrix},\,\begin{bmatrix}3\\3\\-1\\3\end{bmatrix},\,\begin{bmatrix}7\\3\\-6\\4\end{bmatrix}\right\}$$

Solution [143]

C22 Contributed by Robert Beezer

$$\left\{\begin{bmatrix}1\\5\\1\end{bmatrix},\,\begin{bmatrix}6\\-1\\2\end{bmatrix},\,\begin{bmatrix}9\\-3\\8\end{bmatrix},\,\begin{bmatrix}2\\8\\-1\end{bmatrix},\,\begin{bmatrix}3\\-2\\0\end{bmatrix}\right\}$$

Solution [143]

M50 Contributed by Robert Beezer

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 $W = \mathcal{S}p(T)$.

$$W = \mathcal{S}p(\{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\mathbf{v}_4,\,\mathbf{v}_5\}) = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\1\\1\end{bmatrix},\,\begin{bmatrix}-1\\-1\\1\end{bmatrix},\,\begin{bmatrix}1\\2\\3\end{bmatrix},\,\begin{bmatrix}3\\1\\3\end{bmatrix},\,\begin{bmatrix}0\\1\\-3\end{bmatrix}\right\}\right)$$

Solution [143]

T20 Contributed by Robert Beezer

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. Solution [144]

Subsection SOL
Solutions

C20 Exercise [141] Contributed by Robert Beezer
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,

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{bmatrix}$$

the 3×3 identity matrix. So by Theorem NSME2 [137] the original matrix is nonsingular and its columns are therefore a linearly independent set.

C21 Exercise [141] Contributed by Robert Beezer
Theorem LIVRN [133] says we can answer this question by putting theses vectors into a matrix as columns and row-reducing. Doing this we obtain,

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \end{bmatrix}$$

With $n = 3$ (3 vectors, 3 columns) and $r = 3$ (3 leading 1's) we have $n = r$ and the corollary says the vectors are linearly independent.

C22 Exercise [141] Contributed by Robert Beezer
Five vectors from $\mathbb{C}^3$. Theorem MVSLD [133] says the set is linearly dependent. Boom.

M50 Exercise [141] Contributed by Robert Beezer
We want to 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 [121] and Example RSC5 [134]. 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 sytem 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

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

From this we 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 [101] the choice of free variables below lead to solutions and linear combinations, which are then rearranged.

$$x_4 = 1, x_5 = 0 \quad\Rightarrow\quad (-2)\mathbf{v}_1 + (-1)\mathbf{v}_2 + (0)\mathbf{v}_3 + (1)\mathbf{v}_4 + (0)\mathbf{v}_5 = \mathbf{0} \quad\Rightarrow\quad \mathbf{v}_4 = 2\mathbf{v}_1 + \mathbf{v}_2$$
$$x_4 = 0, x_5 = 1 \quad\Rightarrow\quad (1)\mathbf{v}_1 + (2)\mathbf{v}_2 + (0)\mathbf{v}_3 + (0)\mathbf{v}_4 + (1)\mathbf{v}_5 = \mathbf{0} \quad\Rightarrow\quad \mathbf{v}_5 = -\mathbf{v}_1 - 2\mathbf{v}_2$$

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. Thus

$$W = \mathcal{S}p(\{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3\}) = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\1\\1\end{bmatrix},\,\begin{bmatrix}-1\\-1\\1\end{bmatrix},\,\begin{bmatrix}1\\2\\3\end{bmatrix}\right\}\right)$$

There are other answers to this question, but notice that any nontrvial 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.

Though it was not requested in the problem, notice that $\{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3\}$ is a linearly independent set, so we cannot make the set any smaller and still span W.

T20 Exercise [141] Contributed by Robert Beezer

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 [129]) and write a relation of linear dependence (Definition RLDCV [129]),

$$\alpha_1\left(\mathbf{v}_1\right) + \alpha_2\left(\mathbf{v}_1 + \mathbf{v}_2\right) + \alpha_3\left(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3\right) + \alpha_4\left(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4\right) = \mathbf{0}$$

Using the distributive and commutative properties of vector addition and scalar multiplication (Theorem VSPCM [94]) this equation can be rearranged as

$$\left(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4\right)\mathbf{v}_1 + \left(\alpha_2 + \alpha_3 + \alpha_4\right)\mathbf{v}_2 + \left(\alpha_3 + \alpha_4\right)\mathbf{v}_3 + \left(\alpha_4\right)\mathbf{v}_4 = \mathbf{0}$$

However, this is a relation of linear dependence (Definition RLDCV [129]) on a linearly independent set, $\{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\mathbf{v}_4\}$ (this was our lone hypothesis). By the definition of linear independence (Definition LICV [129]) the scalars must all be zero. This is the homogeneous system of equations,

$$\begin{aligned}\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0\\ \alpha_2 + \alpha_3 + \alpha_4 &= 0\\ \alpha_3 + \alpha_4 &= 0\\ \alpha_4 &= 0\end{aligned}$$

Row-reducing the coefficient matrix of this system (or backsolving) gives the conclusion

$$\alpha_1 = 0 \qquad \alpha_2 = 0 \qquad \alpha_3 = 0 \qquad \alpha_4 = 0$$

This means, by Definition LICV [129], 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.

Section O
Orthogonality

In this section we define a couple more operations with vectors, and prove a couple of theorems. These definitions and results are not central to what follows, but we will make use of them frequently throughout the remainder of the course on various occasions. 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 [547].

First, we 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 [91] and Definition CVSM [92]). We can also extend the idea of the conjugate to vectors.

Definition CCV
Conjugate of a Column Vector
Suppose that

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_m \end{bmatrix}$$

is a vector from $\mathbb{C}^m$. Then the conjugate of the vector is defined as

$$\overline{\mathbf{u}} = \begin{bmatrix} \overline{u}_1 \\ \overline{u}_2 \\ \overline{u}_3 \\ \vdots \\ \overline{u}_m \end{bmatrix} \qquad \triangle$$

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 CCRVA
Complex 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 Apply the definition of vector addition (Definition CVA [91]) and the definition of the conjugate of a vector (Definition CCV [145]), and in each component apply the similar property for complex numbers (Theorem CCRA [548]). ■

Theorem CCRSM
Complex Conjugation Respects Scalar Multiplication
Suppose $\mathbf{x}$ is a vector from $\mathbb{C}^m$, and $\alpha \in \mathbb{C}$ is a scalar. Then

$$\overline{\alpha\mathbf{x}} = \overline{\alpha}\,\overline{\mathbf{x}}$$ □

Proof Apply the definition of scalar multiplication (Theorem CVSM [250]) and the definition of the conjugate of a vector (Definition CCV [145]), and in each component apply the similar property for complex numbers (Theorem CCRM [548]). ■

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

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_m \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix}$$

the **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is the scalar quantity in $\mathbb{C}$,

$$\langle \mathbf{u},\, \mathbf{v} \rangle = u_1\overline{v_1} + u_2\overline{v_2} + u_3\overline{v_3} + \cdots + u_m\overline{v_m} = \sum_{i=1}^{m} u_i\overline{v_i}$$ △

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
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

$$\begin{aligned} \langle \mathbf{u},\, \mathbf{v} \rangle &= (2+3i)(\overline{1+2i}) + (5+2i)(\overline{-4+5i}) + (3+i)(\overline{0+5i}) \\ &= (2+3i)(1-2i) + (5+2i)(-4-5i) + (3+i)(0-5i) \\ &= (8-i) + (-10-33i) + (5+15i) \\ &= 3-19i \end{aligned}$$

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

$$\langle \mathbf{w},\, \mathbf{x} \rangle = 2(\overline{3}) + 4(\overline{1}) + (-3)(\overline{0}) + 2(\overline{-1}) + 8(\overline{-2}) = 2(3) + 4(1) + (-3)0 + 2(-1) + 8(-2) = -8.$$

⊚

In the case where the entries of our vectors are all real numbers (as in the second part of Example CSIP [147]), 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 $\mathbb{R}^n$).

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 2 and you can prove part 1.

$$\begin{aligned}
\langle \mathbf{u},\,\mathbf{v}+\mathbf{w}\rangle &= \sum_{i=1}^{m} u_i(\overline{v_i + w_i}) && \text{Definition IP [146]}\\
&= \sum_{i=1}^{m} u_i(\overline{v_i} + \overline{w_i}) && \text{Theorem CCRA [548]}\\
&= \sum_{i=1}^{m} u_i\overline{v_i} + u_i\overline{w_i} && \text{Distributivity}\\
&= \sum_{i=1}^{m} u_i\overline{v_i} + \sum_{i=1}^{m} u_i\overline{w_i} && \text{Commutativity}\\
&= \langle \mathbf{u},\,\mathbf{v}\rangle + \langle \mathbf{u},\,\mathbf{w}\rangle && \text{Definition IP [146]} \quad \blacksquare
\end{aligned}$$

Theorem IPSM
Inner Product and Scalar Multiplication
Suppose $\mathbf{u},\,\mathbf{v} \in \mathbb{C}^m$ and $\alpha \in \mathbb{C}$. Then

1. $\langle \alpha\mathbf{u},\,\mathbf{v}\rangle = \alpha\,\langle \mathbf{u},\,\mathbf{v}\rangle$
2. $\langle \mathbf{u},\,\alpha\mathbf{v}\rangle = \overline{\alpha}\,\langle \mathbf{u},\,\mathbf{v}\rangle$ □

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 2 and you can prove part 1.

$$\begin{aligned}
\langle \mathbf{u},\,\alpha\mathbf{v}\rangle &= \sum_{i=1}^{m} u_i(\overline{\alpha v_i}) && \text{Definition IP [146]}\\
&= \sum_{i=1}^{m} u_i(\overline{\alpha}\,\overline{v_i}) && \text{Theorem CCRM [548]}\\
&= \overline{\alpha}\sum_{i=1}^{m} u_i\overline{v_i} && \text{Distributivity}\\
&= \overline{\alpha}\,\langle \mathbf{u},\,\mathbf{v}\rangle && \text{Definition IP [146]} \quad \blacksquare
\end{aligned}$$

Theorem IPAC
Inner Product is Anti-Commutative
Suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{C}^m$. Then $\langle \mathbf{u},\,\mathbf{v}\rangle = \overline{\langle \mathbf{v},\,\mathbf{u}\rangle}$. □

Proof

$$
\begin{aligned}
\langle \mathbf{u},\, \mathbf{v}\rangle &= u_1\overline{v_1} + u_2\overline{v_2} + u_3\overline{v_3} + \cdots + u_m\overline{v_m}\\
&= \overline{\overline{u_1}}\,\overline{v_1} + \overline{\overline{u_2}}\,\overline{v_2} + \overline{\overline{u_3}}\,\overline{v_3} + \cdots + \overline{\overline{u_m}}\,\overline{v_m}\\
&= \overline{\overline{u_1}v_1} + \overline{\overline{u_2}v_2} + \overline{\overline{u_3}v_3} + \cdots + \overline{\overline{u_m}v_m}\\
&= \overline{\overline{u_1}v_1 + \overline{u_2}v_2 + \overline{u_3}v_3 + \cdots + \overline{u_m}v_m}\\
&= \overline{v_1\overline{u_1} + v_2\overline{u_2} + v_3\overline{u_3} + \cdots + v_m\overline{u_m}}\\
&= \overline{\langle \mathbf{v},\, \mathbf{u}\rangle}
\end{aligned}
$$

■

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.

Definition NV
Norm of a Vector
The **norm** of the vector

$$\mathbf{u} = \begin{bmatrix} u_1\\ u_2\\ u_3\\ \vdots\\ u_m \end{bmatrix}$$

is the scalar quantity in $\mathbb{C}^m$

$$\|\mathbf{u}\| = \sqrt{|u_1|^2 + |u_2|^2 + |u_3|^2 + \cdots + |u_m|^2} = \sqrt{\sum_{i=1}^{m} |u_i|^2}$$

△

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.

Theorem IPN
Inner Products and Norms
Suppose that $\mathbf{u}$ is a vector in $\mathbb{C}^m$. Then $\|\mathbf{u}\|^2 = \langle\mathbf{u},\,\mathbf{u}\rangle$. □

Proof

$$\begin{aligned}
\|\mathbf{u}\|^2 &= \left(\sqrt{|u_1|^2 + |u_2|^2 + |u_3|^2 + \cdots + |u_m|^2}\right)^2 \\
&= |u_1|^2 + |u_2|^2 + |u_3|^2 + \cdots + |u_m|^2 \\
&= u_1\overline{u_1} + u_2\overline{u_2} + u_3\overline{u_3} + \cdots + u_m\overline{u_m} \\
&= \langle\mathbf{u},\,\mathbf{u}\rangle
\end{aligned}$$ ■

When our vectors have entries only from the real numbers Theorem IPN [150] 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 $\mathbf{u}$ is a vector in $\mathbb{C}^m$. Then $\langle\mathbf{u},\,\mathbf{u}\rangle \geq 0$ with equality if and only if $\mathbf{u} = \mathbf{0}$.□

Proof From the proof of Theorem IPN [150] we see that

$$\langle\mathbf{u},\,\mathbf{u}\rangle = |u_1|^2 + |u_2|^2 + |u_3|^2 + \cdots + |u_m|^2$$

Since each modulus 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

$$0 = \langle \mathbf{u}, \mathbf{u} \rangle = |u_1|^2 + |u_2|^2 + |u_3|^2 + \cdots + |u_m|^2$$

Now we have a sum of squares equaling zero, so each term must be zero. Then by similar logic, $|u_i| = 0$ will imply that $u_i = 0$, 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. ■

The conditions of Theorem PIP [150] 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, $\mathbf{u}$ and $\mathbf{v}$, 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

$$\mathbf{u} = \begin{bmatrix} 2+3i \\ 4-2i \\ 1+i \\ 1+i \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 1-i \\ 2+3i \\ 4-6i \\ 1 \end{bmatrix}$$

are orthogonal since

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= (2+3i)(1+i) + (4-2i)(2-3i) + (1+i)(4+6i) + (1+i)(1) \\ &= (-1+5i) + (2-16i) + (-2+10i) + (1+i) \\ &= 0 + 0i. \end{aligned}$$

⊚

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 $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_n\}$ is a set of vectors from $\mathbb{C}^m$. Then S is **orthogonal** if every pair of different vectors from S is orthogonal, that is $\langle \mathbf{u}_i,\, \mathbf{u}_j \rangle = 0$ whenever $i \neq j$. △

The next example is trivial in some respects, but is still worthy of discussion since it is the prototypical orthogonal set.

Example SUVOS
Standard Unit Vectors are an Orthogonal Set
The standard unit vectors are the columns of the identity matrix (Definition SUV [218]). Computing the inner product of two distinct vectors, $\mathbf{e}_i,\, \mathbf{e}_j,\ i \neq j$, gives,

$$\begin{aligned}\langle \mathbf{e}_i,\, \mathbf{e}_j \rangle &= 0\overline{0} + 0\overline{0} + \cdots + 1\overline{0} + \cdots + 0\overline{1} + \cdots + 0\overline{0} + 0\overline{0} \\ &= 0(0) + 0(0) + \cdots + 1(0) + \cdots + 0(1) + \cdots + 0(0) + 0(0) \\ &= 0\end{aligned}$$

⊚

Example AOS
An orthogonal set
The set

$$\{\mathbf{x}_1,\, \mathbf{x}_2,\, \mathbf{x}_3,\, \mathbf{x}_4\} = \left\{ \begin{bmatrix} 1+i \\ 1 \\ 1-i \\ i \end{bmatrix},\, \begin{bmatrix} 1+5i \\ 6+5i \\ -7-i \\ 1-6i \end{bmatrix},\, \begin{bmatrix} -7+34i \\ -8-23i \\ -10+22i \\ 30+13i \end{bmatrix},\, \begin{bmatrix} -2-4i \\ 6+i \\ 4+3i \\ 6-i \end{bmatrix} \right\}$$

is an orthogonal set. Since the inner product is anti-commutative (Theorem IPAC [148]) 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 pairs of different vectors to use in an inner product computation. We'll do two and you can practice your inner products on the other four.

$$\begin{aligned}\langle \mathbf{x}_1,\, \mathbf{x}_3 \rangle &= (1+i)(-7-34i) + (1)(-8+23i) + (1-i)(-10-22i) + (i)(30-13i) \\ &= (27-41i) + (-8+23i) + (-32-12i) + (13+30i) \\ &= 0 + 0i\end{aligned}$$

and

$$\begin{aligned}\langle \mathbf{x}_2,\, \mathbf{x}_4 \rangle &= (1+5i)(-2+4i) + (6+5i)(6-i) + (-7-i)(4-3i) + (1-6i)(6+i) \\ &= (-22-6i) + (41+24i) + (-31+17i) + (12-35i) \\ &= 0 + 0i\end{aligned}$$

⊚

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 $S = \{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_n\}$ is an orthogonal set of nonzero vectors. Then S is linearly independent. □

Proof To prove linear independence of a set of vectors, we can appeal to the definition (Definition LICV [129]) and begin with a relation of linear dependence (Definition RLDCV [129]),

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n = \mathbf{0}.$$

Then, for every $1 \leq i \leq n$, we have

$$\begin{aligned}
0 &= 0\left\langle \mathbf{u}_i,\,\mathbf{u}_i\right\rangle \\
&= \left\langle 0\mathbf{u}_i,\,\mathbf{u}_i\right\rangle && \text{Theorem IPSM [148]}\\
&= \left\langle \mathbf{0},\,\mathbf{u}_i\right\rangle && \text{Theorem CVSM [250]}\\
&= \left\langle \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n,\,\mathbf{u}_i\right\rangle && \text{Relation of linear dependence}\\
&= \left\langle \alpha_1\mathbf{u}_1,\,\mathbf{u}_i\right\rangle + \left\langle \alpha_2\mathbf{u}_2,\,\mathbf{u}_i\right\rangle + \left\langle \alpha_3\mathbf{u}_3,\,\mathbf{u}_i\right\rangle + \cdots + \left\langle \alpha_n\mathbf{u}_n,\,\mathbf{u}_i\right\rangle && \text{Theorem IPVA [147]}\\
&= \alpha_1\left\langle \mathbf{u}_1,\,\mathbf{u}_i\right\rangle + \alpha_2\left\langle \mathbf{u}_2,\,\mathbf{u}_i\right\rangle + \alpha_3\left\langle \mathbf{u}_3,\,\mathbf{u}_i\right\rangle \\
&\quad + \cdots + \alpha_i\left\langle \mathbf{u}_i,\,\mathbf{u}_i\right\rangle + \cdots + \alpha_n\left\langle \mathbf{u}_n,\,\mathbf{u}_i\right\rangle && \text{Theorem IPSM [148]}\\
&= \alpha_1(0) + \alpha_2(0) + \alpha_3(0) + \cdots + \alpha_i\left\langle \mathbf{u}_i,\,\mathbf{u}_i\right\rangle + \cdots + \alpha_n(0) && \text{Orthogonal set}\\
&= \alpha_i\left\langle \mathbf{u}_i,\,\mathbf{u}_i\right\rangle
\end{aligned}$$

So we have $0 = \alpha_i\left\langle \mathbf{u}_i,\,\mathbf{u_i}\right\rangle$. However, since $\mathbf{u}_i \neq \mathbf{0}$ (the hypothesis said our vectors were nonzero), Theorem PIP [150] says that $\left\langle \mathbf{u}_i,\,\mathbf{u_i}\right\rangle > 0$. So we must conclude that $\alpha_i = 0$ for all $1 \leq i \leq n$. 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, with all the scalars zero. Boom! ■

Subsection GSP
Gram-Schmidt Procedure

TODO: Proof technique on induction.

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 $\mathcal{S}p(S) = \mathcal{S}p(T)$. 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 one.

Theorem GSPCV
Gram-Schmidt Procedure, Column Vectors
Suppose that $S = \{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\ldots,\,\mathbf{v}_p\}$ is a linearly independent set of vectors in $\mathbb{C}^m$.

Define the vectors $\mathbf{u}_i$, $1 \leq i \leq p$ by

$$\mathbf{u}_i = \mathbf{v}_i - \frac{\langle \mathbf{v}_i,\, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_i,\, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2,\, \mathbf{u}_2 \rangle}\mathbf{u}_2 - \frac{\langle \mathbf{v}_i,\, \mathbf{u}_3 \rangle}{\langle \mathbf{u}_3,\, \mathbf{u}_3 \rangle}\mathbf{u}_3 - \cdots - \frac{\langle \mathbf{v}_i,\, \mathbf{u}_{i-1} \rangle}{\langle \mathbf{u}_{i-1},\, \mathbf{u}_{i-1} \rangle}\mathbf{u}_{i-1}$$

Then if $T = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_p\}$, then T is an orthogonal set of non-zero vectors, and $\mathcal{S}p(T) = \mathcal{S}p(S)$. □

Proof We will prove the result by using induction on p. 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, $\mathcal{S}p(S) = \mathcal{S}p(T)$. Equally trivial, T is an orthogonal set. If $\mathbf{u}_1 = \mathbf{0}$, then S would be a linearly dependent set, a contradiction.

Now 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 $\mathcal{S}p(S') = \mathcal{S}p(T')$. Define

$$\mathbf{u}_p = \mathbf{v}_p - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2,\, \mathbf{u}_2 \rangle}\mathbf{u}_2 - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_3 \rangle}{\langle \mathbf{u}_3,\, \mathbf{u}_3 \rangle}\mathbf{u}_3 - \cdots - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_{p-1} \rangle}{\langle \mathbf{u}_{p-1},\, \mathbf{u}_{p-1} \rangle}\mathbf{u}_{p-1}$$

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 $\mathcal{S}p(T) = \mathcal{S}p(S)$, by first establishing that $\mathcal{S}p(T) \subseteq \mathcal{S}p(S)$. Suppose $\mathbf{x} \in \mathcal{S}p(T)$, so

$$\mathbf{x} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_p\mathbf{u}_p$$

The term $a_p\mathbf{u}_\mathbf{p}$ is a linear combination of vectors from T' and the vector $\mathbf{v}_\mathbf{p}$, while the remaining terms are a linear combination of vectors from T'. Since $\mathcal{S}p(T') = \mathcal{S}p(S')$, any term that is a multiple of a vector from T' can be rewritten as a linear combinations of vectors from S'. The remaining term $a_p\mathbf{v}_\mathbf{p}$ 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. $\mathbf{x} \in \mathcal{S}p(S)$.

To show that $\mathcal{S}p(S) \subseteq \mathcal{S}p(T)$, begin with $\mathbf{y} \in \mathcal{S}p(S)$, so

$$\mathbf{y} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_p\mathbf{v}_p$$

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 $\mathcal{S}p(S') = \mathcal{S}p(T')$. 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. $\mathbf{y} \in \mathcal{S}p(T)$.

The elements of T' are nonzero, but what about $\mathbf{u}_p$? Suppose to the contrary that

$\mathbf{u}_p = \mathbf{0}$,

$$\mathbf{0} = \mathbf{u}_p = \mathbf{v}_p - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2,\, \mathbf{u}_2 \rangle}\mathbf{u}_2 - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_3 \rangle}{\langle \mathbf{u}_3,\, \mathbf{u}_3 \rangle}\mathbf{u}_3 - \cdots - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_{p-1} \rangle}{\langle \mathbf{u}_{p-1},\, \mathbf{u}_{p-1} \rangle}\mathbf{u}_{p-1}$$

$$\mathbf{v}_p = \frac{\langle \mathbf{v}_p,\, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\, \mathbf{u}_1 \rangle}\mathbf{u}_1 + \frac{\langle \mathbf{v}_p,\, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2,\, \mathbf{u}_2 \rangle}\mathbf{u}_2 + \frac{\langle \mathbf{v}_p,\, \mathbf{u}_3 \rangle}{\langle \mathbf{u}_3,\, \mathbf{u}_3 \rangle}\mathbf{u}_3 + \cdots + \frac{\langle \mathbf{v}_p,\, \mathbf{u}_{p-1} \rangle}{\langle \mathbf{u}_{p-1},\, \mathbf{u}_{p-1} \rangle}\mathbf{u}_{p-1}$$

Since $\mathcal{S}p(S') = \mathcal{S}p(T')$ 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 [134]), 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 orthogonal. We just need to test inner products between $\mathbf{u}_p$ and $\mathbf{u}_i$, for $1 \leq i \leq p-1$. Here we go, using summation notation,

$$\begin{aligned}
\langle \mathbf{u}_p,\, \mathbf{u}_i \rangle &= \left\langle \mathbf{v}_p - \sum_{k=1}^{p-1} \frac{\langle \mathbf{v}_p,\, \mathbf{u}_k \rangle}{\langle \mathbf{u}_k,\, \mathbf{u}_k \rangle}\mathbf{u}_k,\, \mathbf{u}_i \right\rangle \\
&= \langle \mathbf{v}_p,\, \mathbf{u}_i \rangle - \left\langle \sum_{k=1}^{p-1} \frac{\langle \mathbf{v}_p,\, \mathbf{u}_k \rangle}{\langle \mathbf{u}_k,\, \mathbf{u}_k \rangle}\mathbf{u}_k,\, \mathbf{u}_i \right\rangle && \text{Theorem IPVA [147]} \\
&= \langle \mathbf{v}_p,\, \mathbf{u}_i \rangle - \sum_{k=1}^{p-1} \left\langle \frac{\langle \mathbf{v}_p,\, \mathbf{u}_k \rangle}{\langle \mathbf{u}_k,\, \mathbf{u}_k \rangle}\mathbf{u}_k,\, \mathbf{u}_i \right\rangle && \text{Theorem IPVA [147]} \\
&= \langle \mathbf{v}_p,\, \mathbf{u}_i \rangle - \sum_{k=1}^{p-1} \frac{\langle \mathbf{v}_p,\, \mathbf{u}_k \rangle}{\langle \mathbf{u}_k,\, \mathbf{u}_k \rangle} \langle \mathbf{u}_k,\, \mathbf{u}_i \rangle && \text{Theorem IPSM [148]} \\
&= \langle \mathbf{v}_p,\, \mathbf{u}_i \rangle - \frac{\langle \mathbf{v}_p,\, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i,\, \mathbf{u}_i \rangle} \langle \mathbf{u}_i,\, \mathbf{u}_i \rangle - \sum_{k \neq i} \frac{\langle \mathbf{v}_p,\, \mathbf{u}_k \rangle}{\langle \mathbf{u}_k,\, \mathbf{u}_k \rangle}(0) && T' \text{ orthogonal} \\
&= \langle \mathbf{v}_p,\, \mathbf{u}_i \rangle - \langle \mathbf{v}_p,\, \mathbf{u}_i \rangle - \sum_{k \neq i} 0 \\
&= 0
\end{aligned}$$

■

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

$$\mathbf{u}_1 = \mathbf{v_1} = \begin{bmatrix} 1 \\ 1+i \\ 1 \end{bmatrix}$$

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2,\, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\, \mathbf{u}_1 \rangle}\mathbf{u}_1 = \frac{1}{4}\begin{bmatrix} -2-3i \\ 1-i \\ 2+5i \end{bmatrix}$$

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3,\, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1,\, \mathbf{u}_1 \rangle}\mathbf{u}_1 - \frac{\langle \mathbf{v}_3,\, \mathbf{u}_2 \rangle}{\langle \mathbf{u}_2,\, \mathbf{u}_2 \rangle}\mathbf{u}_2 = \frac{1}{11}\begin{bmatrix} -3-i \\ 1+3i \\ -1-i \end{bmatrix}$$

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 $\mathcal{S}p(T) = \mathcal{S}p(S)$ (all by Theorem GSPCV [153]). Of course, as a by-product of orthogonality, the set T is also linearly independent (Theorem OSLI [153]). ⊚

One final definition related to orthogonal vectors.

Definition ONS
OrthoNormal Set
Suppose $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_n\}$ 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 [148]).

Example ONTV
Orthonormal set, three vectors
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 [155] is an orthogonal set. We compute the norm of each vector,

$$\|\mathbf{u}_1\| = 2 \qquad \|\mathbf{u}_2\| = \frac{1}{2}\sqrt{11} \qquad \|\mathbf{u}_3\| = \frac{\sqrt{2}}{\sqrt{11}}$$

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

$$S = \left\{ \begin{bmatrix} 1+i \\ 1 \\ 1-i \\ i \end{bmatrix}, \begin{bmatrix} i \\ 1+i \\ -1 \\ -i \end{bmatrix}, \begin{bmatrix} i \\ -i \\ -1+i \\ 1 \end{bmatrix}, \begin{bmatrix} -1-i \\ i \\ 1 \\ -1 \end{bmatrix} \right\}$$

to an orthogonal set via the Gram-Schmidt Process (Theorem GSPCV [153]) and then scale the vectors to norm 1 to create an orthogonal set. You should get the same set you would if you scaled the orthogonal set of Example AOS [152] to become an orthonormal set. ⊚

Over the course of the next couple of chapters we will discover that orthonormal sets have some very nice properties (in addition to being linearly independent).

M: Matrices

Section MO
Matrix Operations

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, and in this section we will backup and start simple. We start with the definition of an important set.

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. △

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
Matrix Equality
The $m \times n$ matrices

$$A = (a_{ij}) \qquad B = (b_{ij})$$

are **equal**, written $A = B$ provided $a_{ij} = b_{ij}$ for all $1 \leq i \leq m$, $1 \leq j \leq n$. △

So equality of matrices translates to the equality of complex numbers, on an entry-by-entry basis. Notice that we now have our *fourth* 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 top-level 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

$$A = (a_{ij}) \qquad\qquad B = (b_{ij})$$

define the **sum** of A and B to be $A + B = C = (c_{ij})$, where

$$c_{ij} = a_{ij} + b_{ij}, \quad 1 \leq i \leq m,\, 1 \leq j \leq n. \qquad \triangle$$

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.

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

$$A + B = \begin{bmatrix} 2 & -3 & 4 \\ 1 & 0 & -7 \end{bmatrix} + \begin{bmatrix} 6 & 2 & -4 \\ 3 & 5 & 2 \end{bmatrix} = \begin{bmatrix} 2+6 & -3+2 & 4+(-4) \\ 1+3 & 0+5 & -7+2 \end{bmatrix} = \begin{bmatrix} 8 & -1 & 0 \\ 4 & 5 & -5 \end{bmatrix} \quad \circledcirc$$

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 SMM
Scalar Matrix Multiplication
Given the $m \times n$ matrix $A = (a_{ij})$ and the scalar $\alpha \in \mathbb{C}$, the **scalar multiple** of A by α is the matrix $\alpha A = C = (c_{ij})$, where

$$c_{ij} = \alpha a_{ij}, \quad 1 \leq i \leq m,\, 1 \leq j \leq n. \qquad \triangle$$

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

$$\alpha A = 7 \begin{bmatrix} 2 & 8 \\ -3 & 5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 7(2) & 7(8) \\ 7(-3) & 7(5) \\ 7(0) & 7(1) \end{bmatrix} = \begin{bmatrix} 14 & 56 \\ -21 & 35 \\ 0 & 7 \end{bmatrix}$$ ⊚

Its usually straightforward to have a calculator do these computations.

Subsection VSP
Vector Space Properties

To refer to matrices abstractly, we have used notation like $A = (a_{ij})$ to connect the name of a matrix with names for its individual entries. As expressions for matrices become more complicated, we will find this notation more cumbersome. So here's some notation that will help us along the way.

Notation ME
Matrix Entries ($[A]_{ij}$)
For a matrix A, the notation $[A]_{ij}$ will refer to the complex number in row i and column j of A. ⊠

As an example, we could rewrite the defining property for matrix addition as

$$[A + B]_{ij} = [A]_{ij} + [B]_{ij}\,.$$

Be careful with this notation, 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 all the entries at once. You might see some of the motivation for this notation in the definition of matrix equality, Definition ME [159].

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 VSPM
Vector Space Properties of M_{mn}
Suppose that A, B and C are $m \times n$ matrices in M_{mn} and α and β are scalars. Then

1. $A + B \in M_{mn}$ (Additive closure)
2. $\alpha A \in M_{mn}$ (Scalar closure)
3. $A + B = B + A$ (Commutativity)
4. $A + (B + C) = (A + B) + C$ (Associativity of matrix addition)
5. There is a matrix, $\mathcal{O}$, called the **zero matrix**, such that $A + \mathcal{O} = A$ for all $A \in M_{mn}$. (Additive identity)

6. For each matrix $A \in M_{mn}$, there exists a matrix $-A \in M_{mn}$ so that $A+(-A) = \mathcal{O}$. (Additive inverses)

7. $\alpha(\beta A) = (\alpha\beta)A$ (Associativity of scalar multiplication)

8. $\alpha(A + B) = \alpha A + \alpha B$ (Distributivity across matrix addition)

9. $(\alpha + \beta)A = \alpha A + \beta A$ (Distributivity across addition)

10. $1A = A$ (Scalar multiplication with 1) □

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 VSPCM [94] — while the objects here are more complicated, our notation makes the proofs cleaner.

To prove that $(\alpha+\beta)A = \alpha A+\beta A$, we need to establish the equality of two matrices (see Technique GS [20]). Definition ME [159] 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 [161] comes into play. Ready? Here we go.

For *any* i and j, $1 \leq i \leq m$, $1 \leq j \leq n$,

$$[(\alpha + \beta)A]_{ij} = (\alpha + \beta)\,[A]_{ij} = \alpha\,[A]_{ij} + \beta\,[A]_{ij} = [\alpha A]_{ij} + [\beta A]_{ij} = [\alpha A + \beta A]_{ij}\,.$$

A one-liner! 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. (3) There are several plus signs, and several instances of juxtaposition. Identify each one, and state exactly what operation is being represented by each. (4) State the definition or theorem that makes each step of the proof possible. ■

For now, note the similarities between Theorem VSPM [161] about matrices and Theorem VSPCM [94] 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 $\mathcal{O} = \mathcal{O}_{m\times n} = (z_{ij})$ and defined by $z_{ij} = 0$ for all $1 \leq i \leq m$, $1 \leq j \leq n$. Or, equivalently, $[\mathcal{O}]_{ij} = 0$, for all $1 \leq i \leq m$, $1 \leq j \leq n$. △

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

$$\left[A^t\right]_{ij} = [A]_{ji}, \quad 1 \leq i \leq n,\, 1 \leq j \leq m.$$ △

Example TM
Transpose of a 3×4 matrix
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

$$D^t = \begin{bmatrix} 3 & -1 & 0 \\ 7 & 4 & 3 \\ 2 & 2 & -2 \\ -3 & 8 & 5 \end{bmatrix}.$$ ⊚

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
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 [163] 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. But first, a bit more advice about constructing proofs.

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.

The next theorem is a great place to try this technique. ◊

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 [323] that $A = A^t$. So, in particular, A and A^t have the same size. The size of A^t is $n \times m$, so from $m \times n = n \times m$, we conclude that $m = n$, and hence A must be square. ■

We finish this section with two easy theorems, but they illustrate the interplay of our three new operations, our new notation, and the techniques used to prove matrix equalities.

Theorem TASM
Transposes, Addition, Scalar Multiplication
Suppose that A and B are $m \times n$ matrices. Then

1. $(A + B)^t = A^t + B^t$
2. $(\alpha A)^t = \alpha A^t$ □

Proof Each statement to be proved is an equality of matrices, so we work entry-by-entry and use Definition ME [159]. Think carefully about the objects involved here, the many uses of the plus sign and juxtaposition, and the justification for each step.

$$\left[(A + B)^t\right]_{ij} = [A + B]_{ji} = [A]_{ji} + [B]_{ji} = \left[A^t\right]_{ij} + \left[B^t\right]_{ij} = \left[A^t + B^t\right]_{ij}$$

and

$$\left[(\alpha A)^t\right]_{ij} = [\alpha A]_{ji} = \alpha\,[A]_{ji} = \alpha\left[A^t\right]_{ij} = \left[\alpha A^t\right]_{ij}$$ ■

Theorem TT
Transpose of a Transpose
Suppose that A is an $m \times n$ matrix. Then $\left(A^t\right)^t = A$. □

Proof We again want to prove an equality of matrices, so we work entry-by-entry and use Definition ME [159].

$$\begin{aligned}\left[\left(A^t\right)^t\right]_{ij} &= \left[A^t\right]_{ji}\\ &= [A]_{ij}\end{aligned}$$
■

Subsection MCC
Matrices and Complex Conjugation

As we did with vectors (Definition CCV [145]), 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

$$\left[\overline{A}\right]_{ij} = \overline{[A]_{ij}}$$
△

TODO: An example, and two theorems (interaction with matrix addition and scalar multiplication) belong here.

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 [161] reminds you of what previous theorem? How strong is the similarity?

3. Compute the transpose of the matrix below.

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

Section RM
Range of a Matrix

Theorem SLSLC [101] 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 RM
Range 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 **range** of A, written $\mathcal{R}(A)$, is the subset of $\mathbb{C}^m$ containing all linear combinations of the columns of A,

$$\mathcal{R}(A) = \mathcal{S}p(\{\mathbf{A}_1,\, \mathbf{A}_2,\, \mathbf{A}_3,\, \ldots,\, \mathbf{A}_n\}) \qquad \triangle$$

Subsection RSE
Range 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 RMCS
Range of a matrix and consistent systems
Archetype D [487] and Archetype E [491] are linear systems of equations, with an identical 3×4 coefficient matrix, which we call A here. However, Archetype D [487] is consistent, while Archetype E [491] is not. We can explain this distinction with the range of the matrix A.

The column vector of constants, $\mathbf{b}$, in Archetype D [487] 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 [101], we can summarize this solution as a linear combination of the columns of A that equals $\mathbf{b}$,

$$7\begin{bmatrix}2\\-3\\1\end{bmatrix}+8\begin{bmatrix}1\\4\\1\end{bmatrix}+1\begin{bmatrix}7\\-5\\4\end{bmatrix}+3\begin{bmatrix}-7\\-6\\-5\end{bmatrix}=\begin{bmatrix}8\\-12\\4\end{bmatrix}=\mathbf{b}.$$

This equation says that $\mathbf{b}$ is a linear combination of the columns of A, and then by Definition RM [167], we can say that $\mathbf{b}\in\mathcal{R}(A)$.

On the other hand, Archetype E [491] 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 $\mathbf{c}\not\in\mathcal{R}(A)$, for if it were, then it would equal a linear combination of the columns of A and Theorem SLSLC [101] 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 range. This is the content of the next theorem, and since it is an equivalence, it provides an alternate view of the range.

Theorem RCS
Range and Consistent Systems
Suppose A is an $m\times n$ matrix and $\mathbf{b}$ is a vector of size m. Then $\mathbf{b}\in\mathcal{R}(A)$ if and only if $\mathcal{LS}(A,\,\mathbf{b})$ is consistent. □

Proof (⇒) Suppose $\mathbf{b}\in\mathcal{R}(A)$. Then we can write $\mathbf{b}$ as some linear combination of the columns of A. By Theorem SLSLC [101] we can use the scalars from this linear combination to form a solution to $\mathcal{LS}(A,\,\mathbf{b})$, so this system is consistent.

(⇐) If $\mathcal{LS}(A,\,\mathbf{b})$ is consistent, a solution may be used to write $\mathbf{b}$ as a linear combination of the columns of A. This qualifies $\mathbf{b}$ for membership in $\mathcal{R}(A)$. ■

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 range of A. Or equivalently, it tells us that the range 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.

Given a vector $\mathbf{b}$ and a matrix A it is now very mechanical to test if $\mathbf{b}\in\mathcal{R}(A)$. Form the linear system $\mathcal{LS}(A,\,\mathbf{b})$, row-reduce the augmented matrix, $[A\,|\,\mathbf{b}]$, and test for consistency with Theorem RCLS [56].

Subsection RSOC
Range spanned by original columns

So we have a foolproof, automated procedure for determining membership in $\mathcal{R}(A)$. While this works just fine a vector at a time, we would like to have a more useful description of the set $\mathcal{R}(A)$ as a whole. The next example will preview the first of two fundamental results about the range of a matrix.

Example COC
Casting out columns, Archetype I
Archetype I [510] is a system of linear equations with $m = 4$ equations in $n = 7$ variables. Let I denote the 4×7 coefficient matrix from this system, and consider the range of I, $\mathcal{R}(I)$. By the definition, we have

$$\begin{aligned}\mathcal{R}(I) =&\mathcal{S}p(\{\mathbf{I}_1,\,\mathbf{I}_2,\,\mathbf{I}_3,\,\mathbf{I}_4,\,\mathbf{I}_5,\,\mathbf{I}_6,\,\mathbf{I}_7\})\\ =&\mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\0\\-1\end{bmatrix},\,\begin{bmatrix}4\\8\\0\\-4\end{bmatrix},\,\begin{bmatrix}0\\-1\\2\\2\end{bmatrix},\,\begin{bmatrix}-1\\3\\-3\\4\end{bmatrix},\,\begin{bmatrix}0\\9\\-4\\8\end{bmatrix},\,\begin{bmatrix}7\\-13\\12\\-31\end{bmatrix},\,\begin{bmatrix}-9\\7\\-8\\37\end{bmatrix}\right\}\right)\end{aligned}$$

The set of columns of I is obviously linearly dependent, since we have $n = 7$ vectors from $\mathbb{C}^4$ (see Theorem MVSLD [133]). So we can slim down this set some, and still create the range as the span of a set. The row-reduced form for I is the matrix

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

so we can easily create solutions to the homogeneous system $\mathcal{LS}(I,\,\mathbf{0})$ 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 I. These will allow us to solve for one column vector as a linear combination of some others, in the spirit of Theorem DLDS [134], and remove that vector from the set. We'll set about this task methodically. Set the free variable x_2 to one, and set the other free variables to zero. Then a solution to $\mathcal{LS}(I,\,\mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix}-4\\1\\0\\0\\0\\0\\0\end{bmatrix}$$

which can be used to create the linear combination

$$(-4)\mathbf{I}_1 + 1\mathbf{I}_2 + 0\mathbf{I}_3 + 0\mathbf{I}_4 + 0\mathbf{I}_5 + 0\mathbf{I}_6 + 0\mathbf{I}_7 = \mathbf{0}$$

This can then be arranged and solved for $\mathbf{I}_2$, resulting in $\mathbf{I}_2$ expressed as a linear combination of $\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4\}$,

$$\mathbf{I}_2 = 4\mathbf{I}_1 + 0\mathbf{I}_3 + 0\mathbf{I}_4$$

This means that $\mathbf{I}_2$ is surplus, and we can create $\mathcal{R}(I)$ just as well with a smaller set with this vector removed,

$$\mathcal{R}(I) = \mathcal{S}p(\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4,\,\mathbf{I}_5,\,\mathbf{I}_6,\,\mathbf{I}_7\})$$

Set the free variable x_5 to one, and set the other free variables to zero. Then a solution to $\mathcal{LS}(I,\,\mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} -2 \\ 0 \\ -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

which can be used to create the linear combination

$$(-2)\mathbf{I}_1 + 0\mathbf{I}_2 + (-1)\mathbf{I}_3 + (-2)\mathbf{I}_4 + 1\mathbf{I}_5 + 0\mathbf{I}_6 + 0\mathbf{I}_7 = \mathbf{0}$$

This can then be arranged and solved for $\mathbf{I}_5$, resulting in $\mathbf{I}_5$ expressed as a linear combination of $\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4\}$,

$$\mathbf{I}_5 = 2\mathbf{I}_1 + 1\mathbf{I}_3 + 2\mathbf{I}_4$$

This means that $\mathbf{I}_5$ is surplus, and we can create $\mathcal{R}(I)$ just as well with a smaller set with this vector removed,

$$\mathcal{R}(I) = \mathcal{S}p(\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4,\,\mathbf{I}_6,\,\mathbf{I}_7\})$$

Do it again, set the free variable x_6 to one, and set the other free variables to zero. Then a solution to $\mathcal{LS}(I,\,\mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} -1 \\ 0 \\ 3 \\ 6 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

which can be used to create the linear combination

$$(-1)\mathbf{I}_1 + 0\mathbf{I}_2 + 3\mathbf{I}_3 + 6\mathbf{I}_4 + 0\mathbf{I}_5 + 1\mathbf{I}_6 + 0\mathbf{I}_7 = \mathbf{0}$$

This can then be arranged and solved for $\mathbf{I}_6$, resulting in $\mathbf{I}_6$ expressed as a linear combination of $\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4\}$,

$$\mathbf{I}_6 = 1\mathbf{I}_1 + (-3)\mathbf{I}_3 + (-6)\mathbf{I}_4$$

This means that $\mathbf{I}_6$ is surplus, and we can create $\mathcal{R}(I)$ just as well with a smaller set with this vector removed,

$$\mathcal{R}(I) = \mathcal{S}p(\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4,\,\mathbf{I}_7\})$$

Set the free variable x_7 to one, and set the other free variables to zero. Then a solution to $\mathcal{LS}(I,\,\mathbf{0})$ is

$$\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ -5 \\ -6 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

which can be used to create the linear combination

$$3\mathbf{I}_1 + 0\mathbf{I}_2 + (-5)\mathbf{I}_3 + (-6)\mathbf{I}_4 + 0\mathbf{I}_5 + 0\mathbf{I}_6 + 1\mathbf{I}_7 = \mathbf{0}$$

This can then be arranged and solved for $\mathbf{I}_7$, resulting in $\mathbf{I}_7$ expressed as a linear combination of $\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4\}$,

$$\mathbf{I}_7 = (-3)\mathbf{I}_1 + 5\mathbf{I}_3 + 6\mathbf{I}_4$$

This means that $\mathbf{I}_7$ is surplus, and we can create $\mathcal{R}(I)$ just as well with a smaller set with this vector removed,

$$\mathcal{R}(I) = \mathcal{S}p(\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4\})$$

You might think we could keep this up, but we have run out of free variables. And not coincidentally, the set $\{\mathbf{I}_1,\,\mathbf{I}_3,\,\mathbf{I}_4\}$ is linearly independent (check this!). Hopefully it is clear how each free variable was used to eliminate the corresponding column from the set used to span the range, for this will be the essence of the proof of the next theorem. See if you can mimic this example using Archetype J [515]. 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 [510]. Since the system $\mathcal{LS}(I,\,\mathbf{b})$ is consistent, we know by Theorem RCS [168] that $\mathbf{b} \in \mathcal{R}(I)$. This means that $\mathbf{b}$ must be a linear combination of just $\mathbf{I}_1$, $\mathbf{I}_3$, $\mathbf{I}_4$. Can you find this linear combination? Did you notice that there is just a single (unique) answer? Hmmmm. ⊚

We will now formalize the previous example, which will make it trivial to determine a linearly independent set of vectors that will span the range of a matrix. However, the connections made in the last example are worth working through the example (and Archetype J [515]) carefully before employing the theorem.

Theorem BROC
Basis of the Range with Original Columns
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 $S = \{\mathbf{A}_{d_1}, \mathbf{A}_{d_2}, \mathbf{A}_{d_3}, \ldots, \mathbf{A}_{d_r}\}$. Then

1. $\mathcal{R}(A) = \mathcal{S}p(S)$.
2. S is a linearly independent set. □

Proof We have two conclusions stemming from the same hypothesis. We'll prove the first conclusion first. By definition

$$\mathcal{R}(A) = \mathcal{S}p(\{\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \ldots, \mathbf{A}_n\}).$$

We will find relations of linear dependence on this set of column vectors, each one involving a column corresponding to a free variable along with the columns corresponding to the dependent variables. By expressing the free variable column as a linear combination of the dependent variable columns, we will be able to reduce the set S down to only the set of dependent variable columns while preserving the span.

Let $F = \{f_1, f_2, f_3, \ldots, f_{n-r}\}$ be the sets of column indices where B does not have any leading 1's. For each j, $1 \leq j \leq n - r$ construct the specific solution to the homogeneous system $\mathcal{LS}(A, \mathbf{0})$ given by Theorem VFSLS [104] where the free variables are chosen by the rule

$$\mathbf{x}_{f_i} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

This leads to a solution that is exactly the vector $\mathbf{u}_j$ as defined in Theorem SSNS [121],

$$u_{ij} = \begin{cases} 1 & \text{if } i \in F,\, i = f_j \\ 0 & \text{if } i \in F,\, i \neq f_j \\ -b_{k,f_j} & \text{if } i \in D,\, i = d_k \end{cases}.$$

Then Theorem SLSLC [101] says this solution corresponds to the following relation of linear dependence on the columns of A,

$$(-b_{1f_j})\mathbf{A}_{d_1} + (-b_{2f_j})\mathbf{A}_{d_2} + (-b_{3f_j})\mathbf{A}_{d_3} + \cdots + (-b_{rf_j})\mathbf{A}_{d_r} + (1)\mathbf{A}_{f_j} = \mathbf{0}.$$

This can be rearranged as

$$\mathbf{A}_{f_j} = b_{1f_j}\mathbf{A}_{d_1} + b_{2f_j}\mathbf{A}_{d_2} + b_{3f_j}\mathbf{A}_{d_3} + \cdots + b_{rf_j}\mathbf{A}_{d_r}.$$

This equation can be interpreted to tell us that $\mathbf{A}_{f_j} \in \mathcal{S}p(S)$ for all $1 \leq j \leq n-r$, so $\mathcal{S}p(\{\mathbf{A}_1,\,\mathbf{A}_2,\,\mathbf{A}_3,\,\ldots,\,\mathbf{A}_n\}) \subseteq \mathcal{S}p(S)$. It should be easy to convince yourself (so go ahead and do it!) that the opposite is true, i.e. $\mathcal{S}p(S) \subseteq \mathcal{S}p(\{\mathbf{A}_1,\,\mathbf{A}_2,\,\mathbf{A}_3,\,\ldots,\,\mathbf{A}_n\})$. These two statements about subsets combine (see Technique SE [21]) to give the desired set equality as our conclusion.

Our second conclusion is that S is a linearly independent set. To prove this, we will begin with a relation of linear dependence on S. So suppose there are scalars $\alpha_1,\,\alpha_2,\,\alpha_3,\,\ldots,\,\alpha_r$ so that

$$\alpha_1\mathbf{A}_{d_1} + \alpha_2\mathbf{A}_{d_2} + \alpha_3\mathbf{A}_{d_3} + \cdots + \alpha_r\mathbf{A}_{d_r} = \mathbf{0}.$$

To establish linear independence, we wish to deduce that $\alpha_i = 0$, $1 \leq i \leq r$. This relation of linear dependence allows us to use Theorem SLSLC [101] and construct a solution, $\mathbf{x}$, to the homogeneous system $\mathcal{LS}(A,\,\mathbf{0})$. This vector, $\mathbf{x}$, has α_i in entry d_i, and zeros in every entry at an index in F. This is equivalent then to a solution, $\mathbf{x}$, where each free variable equals zero. What would such a solution look like?

By Theorem VFSLS [104], or from details contained in the proof of Theorem SSNS [121], we see that the only solution to a homogeneous system with the free variables all chosen to be zero is precisely the trivial solution, so $\mathbf{x} = \mathbf{0}$. Since each α_i occurs somewhere as an entry of $\mathbf{x} = \mathbf{0}$, we conclude, as desired, that $\alpha_i = 0$, $1 \leq i \leq r$, and hence the set S is linearly independent. ■

This is a very pleasing result since it gives us a handful of vectors that describe the entire range (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 range (Definition RM [167]) as all linear combinations of the columns of the matrix, and the elements of the set S are still columns of the matrix (we won't be so lucky in the next two constructions of the range).

Procedurally this theorem is very 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. Its still important to study the proof of Theorem BROC [172] and its motivation in Example COC [169]. We'll trot through an example all the same.

Example ROCD
Range with original columns, Archetype D

Let's determine a compact expression for the entire range of the coefficient matrix of the system of equations that is Archetype D [487]. Notice that in Example RMCS [167] we were only determining if individual vectors were in the range or not.

To start with the application of Theorem BROC [172], 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,

$$B = \begin{bmatrix} \boxed{1} & 0 & 3 & -2 \\ 0 & \boxed{1} & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are leading 1's in columns 1 and 2, so $D = \{1, 2\}$. To construct a set that spans $\mathcal{R}(A)$, just grab the columns of A indicated by the set D, so

$$\mathcal{R}(A) = \mathcal{S}p\left(\left\{\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}\right\}\right).$$

That's it.

In Example RMCS [167] we determined that the vector

$$\mathbf{c} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$$

was not in the range of A. Try to write $\mathbf{c}$ as a linear combination of the first two columns of A. What happens?

Also in Example RMCS [167] we determined that the vector

$$\mathbf{b} = \begin{bmatrix} 8 \\ -12 \\ 4 \end{bmatrix}$$

was in the range of A. Try to write $\mathbf{b}$ as a linear combination of the first two columns of A. What happens? Did you find a unique solution to this question? Hmmmm. ⊚

Subsection RNS
The range as null space

We've come to know null spaces quite well, since they are the sets of solutions to homogeneous systems of equations. In this subsection, we will see how to describe the range of a matrix as the null space of a different matrix. Then all of our techniques for studying null spaces can be brought to bear on ranges. As usual, we will begin with an example, and then generalize to a theorem.

Example RNSAD
Range as null space, Archetype D
Begin again with the coefficient matrix of Archetype D [487],

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

and we will describe another approach to finding the range of A.

Theorem RCS [168] says a vector is in the range only if it can be used as the vector of constants for a system of equations with coefficient matrix A and result in a consistent system. So suppose we have an arbitrary vector in the range,

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathcal{R}(A)\,.$$

Then the linear system $\mathcal{LS}(A,\, b)$ will be consistent by Theorem RCS [168]. Let's consider solutions to this system by first creating the augmented matrix

$$[\,A \,|\, \mathbf{b}] = \begin{bmatrix} 2 & 1 & 7 & -7 & b_1 \\ -3 & 4 & -5 & -6 & b_2 \\ 1 & 1 & 4 & -5 & b_3 \end{bmatrix}.$$

To locate solutions we would 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. 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. Here's what you should get:

$$[\,A \,|\, \mathbf{b}] = \begin{bmatrix} \boxed{1} & 0 & 3 & -2 & -\frac{1}{7}b_2 + \frac{4}{7}b_3 \\ 0 & \boxed{1} & 1 & -3 & \frac{1}{7}b_2 + \frac{3}{7}b_3 \\ 0 & 0 & 0 & 0 & b_1 + \frac{1}{7}b_2 - \frac{11}{7}b_3 \end{bmatrix}.$$

Is this a consistent system or not? There's an expression in the last column of the third row, preceded by zeros. Theorem RCLS [56] tells us to look at the leading 1 of the last nonzero row, and see if it is in the final column. Could the expression at the end of the third row be a leading 1 in the last column? The answer is: maybe. It depends on **b**. Some vectors are in the range, some are not. For **b** to be in the range, the system $\mathcal{LS}(A,\, \mathbf{b})$ must be consistent, and the expression in question must not be a leading 1. The only way to prevent it from being a leading 1 is if it is zero, since any nonzero value could be scaled to equal 1 by a row operation. So we have

$$\mathbf{b} \in \mathcal{R}(A) \quad\Longleftrightarrow\quad \mathcal{LS}(A,\, \mathbf{b}) \ \text{ is consistent} \quad\Longleftrightarrow\quad b_1 + \frac{1}{7}b_2 - \frac{11}{7}b_3 = 0.$$

So we have an algebraic description of vectors that are, or are not, in the range. And this description looks like a single linear homogeneous equation in the variables b_1, b_2, b_3. The coefficient matrix of this (simple) homogeneous system has the following coefficient matrix

$$K = \begin{bmatrix} 1 & \frac{1}{7} & -\frac{11}{7} \end{bmatrix}.$$

So we can write that $\mathcal{R}(A) = \mathcal{N}(K)$! Example RMCS [167] has a vector $\mathbf{b} \in \mathcal{R}(A)$ and a vector $\mathbf{c} \notin \mathcal{R}(A)$. Test each of these vectors for membership in $\mathcal{N}(K)$. The four columns of the matrix A are definitely in the range, are they also in $\mathcal{N}(K)$? (The work above tells us these answers shouldn't be surprising, but perhaps doing the computations makes it feel a bit remarkable?).

Theorem BNS [138] tells us how to find a set that spans a null space and that is linearly independent. If we compute this set for K in this example, we find

$$\mathcal{R}(A) = \mathcal{N}(K) = \mathcal{S}p(\{\mathbf{u}_1,\,\mathbf{u}_2\}) = \mathcal{S}p\left(\left\{\begin{bmatrix}-\frac{1}{7}\\1\\0\end{bmatrix},\,\begin{bmatrix}\frac{11}{7}\\0\\1\end{bmatrix}\right\}\right).$$

Can you write these two new vectors ($\mathbf{u}_1$, $\mathbf{u}_2$) each as linear combinations of the columns of A? Uniquely? Can you write each of them as linear combinations of just the first *two* columns of A? Uniquely? Hmmmm.

Doing row operations by hand with variables can be a bit error prone, so lets continue with this example and see if we can improve on it some. Rather than having b_1, b_2, b_3 all moving around in the same column, lets put each in its own column. So if we instead row-reduce

$$\begin{bmatrix}2 & 1 & 7 & -7 & b_1 & 0 & 0\\-3 & 4 & -5 & -6 & 0 & b_2 & 0\\1 & 1 & 4 & -5 & 0 & 0 & b_3\end{bmatrix}$$

we find

$$\begin{bmatrix}\boxed{1} & 0 & 3 & -2 & 0 & -\frac{1}{7}b_2 & \frac{4}{7}b_3\\0 & \boxed{1} & 1 & -3 & 0 & \frac{1}{7}b_2 & \frac{3}{7}b_3\\0 & 0 & 0 & 0 & b_1 & \frac{1}{7}b_2 & -\frac{11}{7}b_3\end{bmatrix}.$$

If we sum the entries of the third row in columns 4, 5 and 6, and set it equal to zero, we get the equation

$$b_1 + \frac{1}{7}b_2 - \frac{11}{7}b_3 = 0$$

which we recognize as the previous condition for membership in the range. Perhaps you can see the row operations reflected in the revised form of the matrix involving the variables. You might also notice that the variables are acting more and more like placeholders (and just getting in the way). Let's try removing them. One more time. Now row-reduce, using a calculator if you like since there are no symbols,

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

to obtain

$$\begin{bmatrix}\boxed{1} & 0 & 3 & -2 & 0 & -\frac{1}{7} & \frac{4}{7}\\0 & \boxed{1} & 1 & -3 & 0 & \frac{1}{7} & \frac{3}{7}\\0 & 0 & 0 & 0 & \boxed{1} & \frac{1}{7} & -\frac{11}{7}\end{bmatrix}.$$

The matrix K from above is now sitting down in the last row, in our "new" columns (5,6,7), and can be read off quickly. Why, we could even program a computer to do it! ⊚

Archetype D [487] is just a tad on the small size for fully motivating the following theorem. In practice, the original matrix may row-reduce to a matrix with several nonzero rows. If we row-reduced an augmented matrix having a variable vector of constants ($\mathbf{b}$), we might find several expressions that need to be zero for the system to be consistent. Notice that it is not enough to make just the last expression zero, as then the one above it would also have to be zero, etc. In this way we typically end up with *several* homogeneous equations prescribing elements of the range, and the coefficient matrix of this system (K) will have several rows.

Here's a theorem based on the preceding example, which will give us another procedure for describing the range of a matrix.

Theorem RNS
Range as a Null Space
Suppose that A is an $m \times n$ matrix. Create the $m \times (n+m)$ matrix M by placing the $m \times m$ identity matrix I_m to the right of the matrix A. Symbolically, $M = [A \mid I_m]$. Let N be a matrix that is row-equivalent to M and in reduced row-echelon form. Suppose there are r leading 1's of N in the first n columns. If $r = m$, then $\mathcal{R}(A) = \mathbb{C}^m$. Otherwise, $r < m$ and let K be the $(m-r) \times m$ matrix formed from the entries of N in the last $m - r$ rows and last m columns. Then

1. K is in reduced row-echelon form.

2. K has no zero rows, or equivalently, K has $m - r$ leading 1's.

3. $\mathcal{R}(A) = \mathcal{N}(K)$. □

Proof Let B denote the $m \times n$ matrix that is the first n columns of N, and let J denote the $m \times m$ matrix that is the final m columns of N. Then the sequence of row operations that convert M into N, will also convert A into B and I_m into J.

When $r = m$, there are m leading 1's in N that occur in the first n columns, so B has no zero rows. Thus, the linear system $\mathcal{LS}(A, \mathbf{b})$ is never inconsistent, no matter which vector is chosen for $\mathbf{b}$. So by Theorem RCS [168], every $\mathbf{b} \in \mathbb{C}^m$ is in $\mathcal{R}(A)$.

Now consider the case when $r < m$. The final $m - r$ rows of B are zero rows since the leading 1's of these rows for N are located in columns $n + 1$ or higher. The final $m - r$ rows of J form the matrix K.

Since N is in reduced row-echelon form, and the first n entries of each of the final $m - r$ rows are zero, K will have leading 1's in an echelon pattern, any zero rows are at the bottom (but we'll soon see that there aren't any), and columns with leading 1's will be otherwise zero. In other words, K is in reduced row-echelon form.

Theorem NSRRI [76] tells us that the matrix I_m is nonsingular, since it is row-equivalent to the identity matrix (by an empty sequence of row operations!). Therefore, it cannot be row-equivalent to a matrix with a zero row. Why not? A square matrix with a zero row is the coefficient matrix of a homogeneous system that has more variables than equations, if we consider the zero row as a "nonexistent" equation. Theorem HMVEI [69] then says the system has infinitely many solutions. In turn this implies that the homogeneous linear system $\mathcal{LS}(I_m, \mathbf{0})$ has infinitely many solutions, implying that I_m is singular, a contradiction. Since K is part of J, and J is row-equivalent to I_m, there can be no zero rows in K. If K has no zero rows, then it must have a leading 1 in each of its $m-r$ rows.

For our third conclusion, begin by supposing that $\mathbf{b}$ is an arbitrary vector in $\mathbb{C}^m$. To the vector $\mathbf{b}$ apply the row operations that convert M to N and call the resulting vector $\mathbf{c}$. Then the linear systems $\mathcal{LS}(A,\,\mathbf{b})$ and $\mathcal{LS}(B,\,\mathbf{c})$ are equivalent. Also, the linear systems $\mathcal{LS}(I_m,\,\mathbf{b})$ and $\mathcal{LS}(J,\,\mathbf{c})$ are equivalent and the unique solution to each is simply $\mathbf{b}$. Finally, let $\mathbf{c}^*$ be the vector of length $m-r$ containing the final $m-r$ entries of $\mathbf{c}$. Then we have the following,

$$\begin{aligned}
\mathbf{b}\in\mathcal{R}(A) &\iff \mathcal{LS}(A,\,\mathbf{b}) \text{ is consistent} && \text{Theorem RCS [168]}\\
&\iff \mathcal{LS}(B,\,\mathbf{c}) \text{ is consistent} && \text{Definition ES [18]}\\
&\iff \mathbf{c}^*=\mathbf{0} && \text{Theorem RCLS [56]}\\
&\iff \mathbf{b} \text{ is a solution to } \mathcal{LS}(K,\,\mathbf{0}) && \mathbf{b} \text{ is unique solution to } \mathcal{LS}(J,\,\mathbf{c})\\
&\iff \mathbf{b}\in\mathcal{N}(K)\,. && \text{Definition NSM [73]}
\end{aligned}$$

Running these equivalences in the two different directions will establish the subset inclusions needed by Technique SE [21] and so we can conclude that $\mathcal{R}(A)=\mathcal{N}(K)$. ■

We've commented that Archetype D [487] was a tad small to fully appreciate this theorem. Let's apply it now to an Archetype where there's a bit more action.

Example RNSAG
Range as null space, Archetype G
Archetype G [501] and Archetype H [505] 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

$$N=\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & \frac{3}{11} & \frac{1}{33}\\ 0 & \boxed{1} & 0 & 0 & 0 & -\frac{2}{11} & \frac{1}{11}\\ 0 & 0 & \boxed{1} & 0 & 0 & 0 & -\frac{1}{3}\\ 0 & 0 & 0 & \boxed{1} & 0 & 1 & -\frac{1}{3}\\ 0 & 0 & 0 & 0 & \boxed{1} & 1 & -1\end{bmatrix}.$$

The first $n = 2$ columns contain $r = 2$ leading 1's, so we extract K from the final $m - r = 3$ rows in the final $m = 5$ columns. Since this matrix is guaranteed to be in reduced row-echelon form, we mark the leading 1's.

$$K = \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & \boxed{1} & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & \boxed{1} & 1 & -1 \end{bmatrix}.$$

Theorem RNS [177] now allows us to conclude that $\mathcal{R}(G) = \mathcal{N}(K)$. But we can do better. Theorem BNS [138] tells us how to find a set that spans a null space and that is linearly independent. The matrix K has 3 nonzero rows and 5 columns, so the homogeneous system $\mathcal{LS}(K, \mathbf{0})$ will have solutions described by two free variables x_4 and x_5 in this case. Applying these results in this example yields,

$$\mathcal{R}(G) = \mathcal{N}(K) = \mathcal{S}p(\{\mathbf{u}_1, \mathbf{u}_2\}) = \mathcal{S}p\left(\left\{\begin{bmatrix} 0 \\ -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 1 \\ 0 \\ 1 \end{bmatrix}\right\}\right).$$

As mentioned earlier, Archetype G [501] is consistent, while Archetype H [505] is inconsistent. See if you can write the two different vectors of constants as linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$. How about the two columns of G? They must be in the range of G also. Are your answers unique? Do you notice anything about the scalars that appear in the linear combinations you are forming? ⊚

Subsection RNSM Range of a Nonsingular Matrix

Let's specialize to square matrices and contrast the ranges of the coefficient matrices in Archetype A [473] and Archetype B [478].

Example RAA
Range of Archetype A
The coefficient matrix in Archetype A [473] is

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

which row-reduces to

$$\begin{bmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{1} & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Columns 1 and 2 have leading 1's, so by Theorem BROC [172] we can write

$$\mathcal{R}(A) = \mathcal{S}p(\{\mathbf{A}_1,\, \mathbf{A}_2\}) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\1\end{bmatrix},\, \begin{bmatrix}-1\\1\\1\end{bmatrix}\right\}\right).$$

We want to show in this example that $\mathcal{R}(A) \neq \mathbb{C}^3$. 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 $\mathbf{b}$ as a linear combination of $\mathbf{A}_1$ and $\mathbf{A}_2$. Try one of these two computations yourself. (Or try both!). Since $\mathbf{b} \notin \mathcal{R}(A)$, the range 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). ⊚

Example RAB
Range of Archetype B

The coefficient matrix in Archetype B [478], call it B here, is known to be nonsingular (see Example NS [76]). By Theorem NSMUS [79], the linear system $\mathcal{LS}(B,\, \mathbf{b})$ has a (unique) solution for every choice of $\mathbf{b}$. Theorem RCS [168] then says that $\mathbf{b} \in \mathcal{R}(B)$ 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 NSMUS [79]. ⊚

Theorem RNSM
Range of a NonSingular Matrix

Suppose A is a square matrix of size n. Then A is nonsingular if and only if $\mathcal{R}(A) = \mathbb{C}^n$. □

Proof ($\Leftarrow$) Suppose A is nonsingular. By Theorem NSMUS [79], the linear system $\mathcal{LS}(A,\, \mathbf{b})$ has a (unique) solution for every choice of $\mathbf{b}$. Theorem RCS [168] then says that $\mathbf{b} \in \mathcal{R}(A)$ for all $\mathbf{b} \in \mathbb{C}^n$. In other words, $\mathcal{R}(A) = \mathbb{C}^n$.

($\Rightarrow$) We'll prove the contrapositive (see Technique CP [55]). Suppose that A is singular. By Theorem NSRRI [76], A will not row-reduce to the identity matrix I_n. So the row-equivalent matrix B of Theorem RNS [177] has $r < n$ nonzero rows and then the matrix K is a nonzero matrix (it has at least one leading 1 in it). By Theorem NSRRI [76], $\mathcal{R}(A) = \mathcal{N}(K)$. If we can find one vector of $\mathbb{C}^n$ that is not in $\mathcal{N}(K)$, then we can conclude that $\mathcal{R}(A) \neq \mathbb{C}^n$, the desired conclusion for the contrapositive.

The matrix K has at least one nonzero entry, suppose it is located in column t. Let $\mathbf{x} \in \mathbb{C}^n$ be a vector of all zeros, except a 1 in entry t. Use this vector to form a linear combination of the columns of K, and the result will be just column t of K, which is nonzero. So by Theorem SLSLC [101], the vector $\mathbf{x}$ cannot be a solution to the homogeneous system $\mathcal{LS}(K,\, \mathbf{0})$, so $\mathbf{x} \notin \mathcal{N}(K)$. ■

With this equivalence for nonsingular matrices we can update our list, Theorem NSME2 [137].

Theorem NSME3
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. The range of A is $\mathbb{C}^n$, $\mathcal{R}(A) = \mathbb{C}^n$. □

Proof Since Theorem RNSM [180] is an equivalence, we can add it to the list in Theorem NSME2 [137]. ■

Subsection READ Reading Questions

1. Write the range of the matrix below as the span of a set of three vectors.

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

2. List three techniques you could use to provide a description of the range of a matrix.
3. Suppose that A is an $n \times n$ nonsingular matrix. What can you say about its range?

Section RSM
Row Space Of a Matrix

The range of a matrix is sometimes called the **column space** since it is formed by taking all possible linear combinations of the columns of the matrix. We can do a similar construction with the rows of a matrix, and that is the topic of this section. This will provide us with even more connections with row operations. However, we are also going to try to parlay our knowledge of the range, so we'll get at the rows of a matrix by working with the columns of the transpose. A side-benefit will be a third way to describe the range of a matrix.

Subsection RSM
Row Space of a Matrix

Definition RSM
Row Space of a Matrix
Suppose A is an $m \times n$ matrix. Then the **row space** of A, $\nabla s(A)$, is the range of A^t, i.e. $\nabla s(A) = \mathcal{R}(A^t)$. △

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. With the row space defined in terms of the range, all of the results of Section RM [167] can be applied to row spaces.

Notice that if A is a rectangular $m \times n$ matrix, then $\mathcal{R}(A) \subseteq \mathbb{C}^m$, while $\nabla s(A) \subseteq \mathbb{C}^n$ 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 $\mathcal{R}(A)$ and $\nabla s(A)$ are subsets of $\mathbb{C}^n$, though usually the sets will not be equal.

Example RSAI
Row space of Archetype I
The coefficient matrix in Archetype I [510] 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,

$$\nabla s(I) = \mathcal{R}\left(I^t\right) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 0 \\ 7 \\ -9 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \\ -1 \\ 3 \\ 9 \\ -13 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \\ -3 \\ -4 \\ 12 \\ -8 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ 2 \\ 4 \\ 8 \\ -31 \\ 37 \end{bmatrix}\right\}\right).$$

However, we can use Theorem BROC [172] to get a slightly better description. First, row-reduce I^t,

$$\begin{bmatrix} \boxed{1} & 0 & 0 & -\frac{31}{7} \\ 0 & \boxed{1} & 0 & \frac{12}{7} \\ 0 & 0 & \boxed{1} & \frac{13}{7} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since there are leading 1's in columns with indices $D = \{1,\, 2,\, 3\}$, the range of I^t can be spanned by just the first three columns of I^t,

$$\nabla s(I) = \mathcal{R}\left(I^t\right) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 4 \\ 0 \\ -1 \\ 0 \\ 7 \\ -9 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \\ -1 \\ 3 \\ 9 \\ -13 \\ 7 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 2 \\ -3 \\ -4 \\ 12 \\ -8 \end{bmatrix}\right\}\right).$$

The technique of Theorem RNS [177] could also have been applied to the matrix I^t, by adjoining the 7×7 identity matrix, I_7 and row-reducing the resulting 11×7 matrix. The 4×7 matrix K that results is

$$K = \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{3}{7} & -\frac{1}{7} & 0 \\ 0 & 1 & 0 & 0 & -\frac{12}{7} & -\frac{4}{7} & 0 \\ 0 & 0 & 1 & 0 & -\frac{3}{7} & -\frac{9}{14} & -\frac{1}{2} \\ 0 & 0 & 0 & 1 & \frac{1}{7} & \frac{13}{28} & \frac{1}{4} \end{bmatrix}.$$

Then $\nabla s(I) = \mathcal{R}(I^t) = \mathcal{N}(K)$, and we could use Theorem BNS [138] to express the null space of K as the span of three vectors, one for each free variable in the homogeneous system $\mathcal{LS}(K, \mathbf{0})$. ⊚

The row space would not be too interesting if it was simply the range 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 $\nabla s(A) = \nabla s(B)$. □

Proof Two matrices are row-equivalent (Definition REM [36]) 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, $\mathcal{R}(A^t) = \mathcal{R}(B^t)$. We will take a generic element of one and show that it is contained in the other.

$$\begin{aligned}
&\beta_1\mathbf{B}_1 + \beta_2\mathbf{B}_2 + \beta_3\mathbf{B}_3 + \cdots + \beta_t\mathbf{B}_t + \cdots + \beta_m\mathbf{B}_m = \\
&\beta_1\mathbf{A}_1 + \beta_2\mathbf{A}_2 + \beta_3\mathbf{A}_3 + \cdots + \beta_t\left(\alpha\mathbf{A}_t\right) + \cdots + \beta_m\mathbf{A}_m = \\
&\beta_1\mathbf{A}_1 + \beta_2\mathbf{A}_2 + \beta_3\mathbf{A}_3 + \cdots + \left(\alpha\beta_t\right)\mathbf{A}_t + \cdots + \beta_m\mathbf{A}_m
\end{aligned}$$

says that $\mathcal{R}(B^t) \subseteq \mathcal{R}(A^t)$. Similarly,

$$\begin{aligned}
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \gamma_t\mathbf{A}_t + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \left(\frac{\gamma_t}{\alpha}\alpha\right)\mathbf{A}_t + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \frac{\gamma_t}{\alpha}\left(\alpha\mathbf{A}_t\right) + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{B}_1 + \gamma_2\mathbf{B}_2 + \gamma_3\mathbf{B}_3 + \cdots + \frac{\gamma_t}{\alpha}\mathbf{B}_t + \cdots + \gamma_m\mathbf{B}_m
\end{aligned}$$

says that $\mathcal{R}(A^t) \subseteq \mathcal{R}(B^t)$. So $\nabla s(A) = \mathcal{R}(A^t) = \mathcal{R}(B^t) = \nabla s(B)$ 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 $i \neq t$.

$$\begin{aligned}
&\beta_1\mathbf{B}_1 + \beta_2\mathbf{B}_2 + \beta_3\mathbf{B}_3 + \cdots + \beta_s\mathbf{B}_s + \cdots + \beta_t\mathbf{B}_t + \cdots + \beta_m\mathbf{B}_m = \\
&\beta_1\mathbf{A}_1 + \beta_2\mathbf{A}_2 + \beta_3\mathbf{A}_3 + \cdots + \beta_s\mathbf{A}_s + \cdots + \beta_t\left(\alpha\mathbf{A}_s + \mathbf{A}_t\right) + \cdots + \beta_m\mathbf{A}_m = \\
&\beta_1\mathbf{A}_1 + \beta_2\mathbf{A}_2 + \beta_3\mathbf{A}_3 + \cdots + \beta_s\mathbf{A}_s + \cdots + \left(\beta_t\alpha\right)\mathbf{A}_s + \beta_t\mathbf{A}_t + \cdots + \beta_m\mathbf{A}_m = \\
&\beta_1\mathbf{A}_1 + \beta_2\mathbf{A}_2 + \beta_3\mathbf{A}_3 + \cdots + \beta_s\mathbf{A}_s + \left(\beta_t\alpha\right)\mathbf{A}_s + \cdots + \beta_t\mathbf{A}_t + \cdots + \beta_m\mathbf{A}_m = \\
&\beta_1\mathbf{A}_1 + \beta_2\mathbf{A}_2 + \beta_3\mathbf{A}_3 + \cdots + \left(\beta_s + \beta_t\alpha\right)\mathbf{A}_s + \cdots + \beta_t\mathbf{A}_t + \cdots + \beta_m\mathbf{A}_m
\end{aligned}$$

says that $\mathcal{R}(B^t) \subseteq \mathcal{R}(A^t)$. Similarly,

$$\begin{aligned}
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \gamma_s\mathbf{A}_s + \cdots + \gamma_t\mathbf{A}_t + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \gamma_s\mathbf{A}_s + \cdots + \left(-\alpha\gamma_t\mathbf{A}_s + \alpha\gamma_t\mathbf{A}_s\right) + \gamma_t\mathbf{A}_t + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \left(-\alpha\gamma_t\mathbf{A}_s\right) + \gamma_s\mathbf{A}_s + \cdots + \left(\alpha\gamma_t\mathbf{A}_s + \gamma_t\mathbf{A}_t\right) + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{A}_1 + \gamma_2\mathbf{A}_2 + \gamma_3\mathbf{A}_3 + \cdots + \left(-\alpha\gamma_t + \gamma_s\right)\mathbf{A}_s + \cdots + \gamma_t\left(\alpha\mathbf{A}_s + \mathbf{A}_t\right) + \cdots + \gamma_m\mathbf{A}_m = \\
&\gamma_1\mathbf{B}_1 + \gamma_2\mathbf{B}_2 + \gamma_3\mathbf{B}_3 + \cdots + \left(-\alpha\gamma_t + \gamma_s\right)\mathbf{B}_s + \cdots + \gamma_t\mathbf{B}_t + \cdots + \gamma_m\mathbf{B}_m
\end{aligned}$$

says that $\mathcal{R}(A^t) \subseteq \mathcal{R}(B^t)$. So $\nabla s(A) = \mathcal{R}(A^t) = \mathcal{R}(B^t) = \nabla s(B)$ when a single row operation of the third type is performed.

So the row space is preserved by each row operation, and hence row spaces of row-equivalent matrices are equal. ■

Example RSREM
Row spaces of two row-equivalent matrices
In Example TREM [36] we saw that the matrices

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

are row-equivalent by demonstrating a sequence of two row operations that converted A into B. Applying Theorem REMRS [185] we can say

$$\nabla s(A) = \mathcal{S}p\left(\left\{\begin{bmatrix} 2 \\ -1 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 6 \end{bmatrix}\right\}\right) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 1 \\ 0 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ -2 \\ -9 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 3 \\ 4 \end{bmatrix}\right\}\right) = \nabla s(B)$$

⊚

Theorem REMRS [185] 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?). 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. $\nabla s(A) = \mathcal{S}p(S)$.
2. S is a linearly independent set. □

Proof From Theorem REMRS [185] we know that $\nabla s(A) = \nabla s(B)$. 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.

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

$$\alpha_1\mathbf{B}_1 + \alpha_2\mathbf{B}_2 + \alpha_3\mathbf{B}_3 + \cdots + \alpha_r\mathbf{B}_r = \mathbf{0}$$

Now consider this equation across entries of the vectors in location d_i, $1 \leq i \leq r$. Since B is in reduced row-echelon form, the entries of column d_i are all zero, except for a (leading) 1 in row i. Considering the column vectors of B^t, the linear combination for entry d_i is

$$\alpha_1(0) + \alpha_2(0) + \alpha_3(0) + \cdots + \alpha_i(1) + \cdots + \alpha_r(0) = 0$$

and from this we conclude that $\alpha_i = 0$ for all $1 \leq i \leq r$, establishing the linear independence of S. ■

Example IAS
Improving a span

Suppose in the course of analyzing a matrix (its range, its null space, its. . .) we encounter the following set of vectors, described by a span

$$X = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\1\\6\\6\end{bmatrix}, \begin{bmatrix}3\\-1\\2\\-1\\6\end{bmatrix}, \begin{bmatrix}1\\-1\\0\\-1\\-2\end{bmatrix}, \begin{bmatrix}-3\\2\\-3\\6\\-10\end{bmatrix}\right\}\right)$$

Let A be the matrix whose rows are the vectors in X, so by design $\nabla s(A) = X$,

$$A = \begin{bmatrix}1 & 2 & 1 & 6 & 6\\ 3 & -1 & 2 & -1 & 6\\ 1 & -1 & 0 & -1 & -2\\ -3 & 2 & -3 & 6 & -10\end{bmatrix}$$

Row-reduce A to form a row-equivalent matrix in reduced row-echelon form,

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

Then Theorem BRS [186] says we can grab the nonzero columns of B^t and write

$$X = \nabla s(A) = \nabla s(B) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\2\\-1\end{bmatrix},\ \begin{bmatrix}0\\1\\0\\3\\1\end{bmatrix},\ \begin{bmatrix}0\\0\\1\\-2\\5\end{bmatrix}\right\}\right)$$

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.* ⊚

Theorem BRS [186] and the techniques of Example IAS [187] will provide yet another description of the range of a matrix. First we state a triviality as a theorem, so we can reference it later

Theorem RMRST
Range of a Matrix is Row Space of Transpose
Suppose A is a matrix. Then $\mathcal{R}(A) = \nabla s(A^t)$. □

Proof Apply Theorem TASM [164] with Definition RSM [183],

$$\nabla s\left(A^t\right) = \mathcal{R}\left(\left(A^t\right)^t\right) = \mathcal{R}(A)\,.$$ ■

So to find yet another expression for the range 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 spanning set. We'll do Archetype I [510], then you do Archetype J [515].

Example RROI
Range from row operations, Archetype I
To find the range of the coefficient matrix of Archetype I [510], 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,

$$\begin{bmatrix} \boxed{1} & 0 & 0 & -\frac{31}{7} \\ 0 & \boxed{1} & 0 & \frac{12}{7} \\ 0 & 0 & \boxed{1} & \frac{13}{7} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now, using Theorem RMRST [188] and Theorem BRS [186]

$$\mathcal{R}(I) = \mathcal{R}s\left(I^t\right) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \\ -\frac{31}{7} \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \frac{12}{7} \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \frac{13}{7} \end{bmatrix}\right\}\right).$$

This is a very nice description of the range. Fewer vectors than the 7 involved in the definition, and the structure of the zeros and ones in the first 3 slots can be used to advantage. For example, Archetype I [510] is presented as 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 RCS [168] tells us that $\mathbf{b} \in \mathcal{R}(I)$. 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 $\mathbf{b}$.

$$\mathbf{b} = \begin{bmatrix} 3 \\ 9 \\ 1 \\ 4 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 0 \\ 0 \\ -\frac{31}{7} \end{bmatrix} + 9\begin{bmatrix} 0 \\ 1 \\ 0 \\ \frac{12}{7} \end{bmatrix} + 1\begin{bmatrix} 0 \\ 0 \\ 1 \\ \frac{13}{7} \end{bmatrix}$$

Can you now rapidly construct several vectors, $\mathbf{b}$, so that $\mathcal{LS}(I,\, \mathbf{b})$ is consistent, and several more so that the system is inconsistent? ⊚

Example COC [169] and Example RROI [188] each describes the range of the coefficient matrix from Archetype I [510] 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 range of this matrix using the null space of the matrix K from Theorem RNS [177] 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 range of a matrix as the span of a linearly independent set. Theorem BROC [172] is sometimes useful since the vectors it specifies are equal to actual columns of the matrix. Theorem RNS [177] tends to create vectors with lots of zeros, and strategically placed 1's near the bottom of the vector. Finally, Theorem BRS [186] and Theorem RMRST [188] combine to create vectors with lots of zeros, and strategically placed 1's near the top of the vector.

Subsection READ
Reading Questions

1. Describe the row space of a matrix in words.

2. Suppose you wished to find the range of a matrix A. What would be the quickest way to find a linearly independent set S so that the range equaled Sp(S)?

3. Is the vector $\begin{bmatrix} 0 \\ 5 \\ 2 \\ 3 \end{bmatrix}$ in the row space of the following matrix?

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

Subsection EXC
Exercises

C30 Contributed by Robert Beezer
Let A be the matrix below, and find the indicated sets by the requested methods.

$$A = \begin{bmatrix} 2 & -1 & 5 & -3 \\ -5 & 3 & -12 & 7 \\ 1 & 1 & 4 & -3 \end{bmatrix}$$

(a) A linearly independent set S so that $\mathcal{R}(A) = \mathcal{S}p(S)$ and S is composed of columns of A.
(b) A linearly independent set S so that $\mathcal{R}(A) = \mathcal{S}p(S)$ 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 $\mathcal{R}(A) = \mathcal{S}p(S)$ 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 $\nabla s(A) = \mathcal{S}p(S)$. Solution [193]

TODO: Construct an example of a matrix A where $\mathcal{R}(A) = \nabla s(A)$.

Subsection SOL Solutions

C30 Exercise [191] Contributed by Robert Beezer

(a) First find a matrix B that is row-equivalent to A and in reduced row-echelon form

$$B = \begin{bmatrix} \boxed{1} & 0 & 3 & -2 \\ 0 & \boxed{1} & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem BROC [172] 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 range of A as the row space of the transpose. 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 [186], ans 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 RNS [177], augment A with the 3×3 identity matrix I_3 and row-reduce to obtain,

$$\begin{bmatrix} 1 & 0 & 3 & -2 & 0 & -\frac{1}{8} & \frac{3}{8} \\ 0 & 1 & 1 & -1 & 0 & \frac{1}{8} & \frac{5}{8} \\ 0 & 0 & 0 & 0 & 1 & \frac{3}{8} & -\frac{1}{8} \end{bmatrix}$$

Then since the first four columns of row 3 are all zeros, we extract

$$K = \begin{bmatrix} \boxed{1} & \frac{3}{8} & -\frac{1}{8} \end{bmatrix}$$

Theorem RNS [177] says that $\mathcal{R}(A) = \mathcal{N}(K)$. We can then use Theorem BNS [138] to construct the desired set S, based on the free variables with indices in $F = \{2, 3\}$ for the homogeneous system $\mathcal{LS}(K, \mathbf{0})$, 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 [186]. Use the row-reduced matrix B from part (a), grab the nonzero rows, and write them as column vectors,

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \end{bmatrix} \right\}$$

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, Theorem SLSLC [101] 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, motivates 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

$$A\mathbf{u} = [\mathbf{A}_1|\mathbf{A}_2|\mathbf{A}_3|\ldots|\mathbf{A}_n]\begin{bmatrix}u_1\\u_2\\u_3\\\vdots\\u_n\end{bmatrix} = u_1\mathbf{A}_1 + u_2\mathbf{A}_2 + u_3\mathbf{A}_3 + \cdots + u_n\mathbf{A}_n \qquad \triangle$$

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. 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

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

Then

$$A\mathbf{u} = 2\begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix} + (-2)\begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix} + 3\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix} + (-1)\begin{bmatrix} 4 \\ -2 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \\ 6 \end{bmatrix}.$$ ⊚

This definition now makes it possible to represent systems of linear equations compactly in terms of an operation.

Theorem SLEMM
Systems of Linear Equations as Matrix Multiplication
Solutions to the linear system $\mathcal{LS}(A, \mathbf{b})$ are the solutions for $\mathbf{x}$ in the vector equation $A\mathbf{x} = \mathbf{b}$. □

Proof This theorem says (not very clearly) 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 (recall Technique SE [21]). Both of these inclusions are easy with the following chain of equivalences,

$$\begin{aligned} &\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \end{bmatrix} \text{ is a solution to } \mathcal{LS}(A, \mathbf{b}) \\ \iff &x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 + \cdots + x_n\mathbf{A}_n = \mathbf{b} && \text{Theorem SLSLC [101]} \\ \iff &\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_i \end{bmatrix} \text{ is a solution to } A\mathbf{x} = \mathbf{b} && \text{Definition MVP [195].} \end{aligned}$$

■

Example MNSLE
Matrix notation for systems of linear equations
Consider the system of linear equations from Example NSLE [72].

$$\begin{aligned} 2x_1 + 4x_2 - 3x_3 + 5x_4 + x_5 &= 9 \\ 3x_1 + x_2 + \quad + x_4 - 3x_5 &= 0 \\ -2x_1 + 7x_2 - 5x_3 + 2x_4 + 2x_5 &= -3 \end{aligned}$$

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 equation $A\mathbf{x} = \mathbf{b}$. ⊚

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,

$$AB = A\left[\mathbf{B}_1|\mathbf{B}_2|\mathbf{B}_3|\ldots|\mathbf{B}_p\right] = \left[A\mathbf{B}_1|A\mathbf{B}_2|A\mathbf{B}_3|\ldots|A\mathbf{B}_p\right].$$ △

Example PTM
Product of two matrices
Set

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

Then

$$AB = \left[A\begin{bmatrix} 1 \\ -1 \\ 1 \\ 6 \\ 1 \end{bmatrix} \middle| A\begin{bmatrix} 6 \\ 4 \\ 1 \\ 4 \\ -2 \end{bmatrix} \middle| A\begin{bmatrix} 2 \\ 3 \\ 2 \\ -1 \\ 3 \end{bmatrix} \middle| A\begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 0 \end{bmatrix} \right] = \begin{bmatrix} 28 & 17 & 20 & 10 \\ 20 & -13 & -3 & -1 \\ -18 & -44 & 12 & -3 \end{bmatrix}.$$ ⊚

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

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

Then we have two square, 2×2 matrices, so Definition MM [197] allows us to multiply them in either order. We find

$$AB = \begin{bmatrix} 19 & 3 \\ 6 & 2 \end{bmatrix} \qquad BA = \begin{bmatrix} 4 & 12 \\ 4 & 17 \end{bmatrix}$$

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? ⊚

Computation Note MM.MMA
Matrix Multiplication (Mathematica)
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:

```
{{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}
```

Understanding the difference between the last two examples will go a long way to explaining how some *Mathematica* constructs work. ⊕

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 the 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 $A = (a_{ij})$ is an $m \times n$ matrix and $B = (b_{ij})$ is an $n \times p$ matrix. Then the entries of $AB = C = (c_{ij})$ are given by

$$[C]_{ij} = c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} = \sum_{k=1}^{n} [A]_{ik} [B]_{kj} \quad \square$$

Proof The value of c_{ij} lies in column j of the product of A and B, and so by Definition MM [197] is the value in location i of the matrix-vector product $A\mathbf{B}_j$. By Definition MVP [195] this matrix-vector product is a linear combination

$$\begin{aligned} A\mathbf{B}_j &= b_{1j}\mathbf{A}_1 + b_{2j}\mathbf{A}_2 + b_{3j}\mathbf{A}_3 + \cdots + b_{nj}\mathbf{A}_n \\ &= b_{1j}\begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{bmatrix} + b_{2j}\begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{bmatrix} + b_{3j}\begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ \vdots \\ a_{m3} \end{bmatrix} + \cdots + b_{nj}\begin{bmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{bmatrix} \end{aligned}$$

We are after the value in location i of this linear combination. Using Definition CVA [91] and Definition CVSM [92] we course through this linear combination in location i to find

$$b_{1j}a_{i1} + b_{2j}a_{i2} + b_{3j}a_{i3} + \cdots + b_{nj}a_{in}.$$

Reversing the order of the products (regular old multiplication *is* commutative) yields the desired expression for c_{ij}. ■

Example PTMEE
Product of two matrices, entry-by-entry
Consider again the two matrices from Example PTM [197]

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

Then suppose we just wanted the entry of AB in the second row, third column:

$$\begin{aligned}[AB]_{23} =&a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} + a_{24}b_{43} + a_{25}b_{53}\\ =&(0)(2) + (-4)(3) + (1)(2) + (2)(-1) + (3)(3) = -3\end{aligned}$$

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 [199], 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:

$$\begin{aligned}[AB]_{31} =&a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} + a_{34}b_{41} + a_{35}b_{51}\\ =&(-5)(1) + (1)(-1) + (2)(1) + (-3)(6) + (4)(1) = -18\end{aligned}$$

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 [197]. 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 [199]. Since this process may take some practice, use your first computation to check your work. ⊚

Theorem EMP [199] is the way most people compute matrix products by hand. It will also be very useful for the theorems we are going to prove shortly. However, the definition is frequently the most useful for its connections with deeper ideas like the null space and range. For example, an alternative (and popular) definition of the range of an $m \times n$ matrix A would be

$$\mathcal{R}(A) = \{\, A\mathbf{x} \mid \mathbf{x} \in \mathbb{C}^n \,\}\,.$$

We recognize this as saying take *all* the matrix vector products possible with the matrix A. By Definition MVP [195] we see that this means take all possible linear combinations of the columns of A — precisely our version of the definition of the range (Definition RM [167]).

Subsection PMM
Properties of Matrix Multiplication

In this subsection, we collect properties of matrix multiplication and its interaction with matrix addition (Definition MA [160]), scalar matrix multiplication (Definition SMM [160]), the identity matrix (Definition IM [76]), the zero matrix (Definition ZM [163]), conjugation (Theorem MMCC [204]) and the transpose (Definition TM [163]). Whew! Here we go. These are great proofs to practice with, so try to concoct the proofs before reading them, they'll get progressively harder 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,

$$\begin{aligned}
\left[A\mathcal{O}_{n\times p}\right]_{ij} &= \sum_{k=1}^{n} [A]_{ik} \left[\mathcal{O}_{n\times p}\right]_{kj} && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} [A]_{ik}\, 0 && \text{Definition ZM [163]}\\
&= \sum_{k=1}^{n} 0 = 0.
\end{aligned}$$

So every entry of the product is the scalar zero, i.e. the result is the zero matrix. ■

Theorem MMIM
Matrix Multiplication and Identity Matrix
Suppose A is an $m \times n$ matrix. Then
1. $AI_n = A$
2. $I_mA = A$ □

Proof Again, we'll prove (1) and leave (2) to you. Entry-by-entry,

$$\begin{aligned}
[AI_n]_{ij} &= \sum_{k=1}^{n} [A]_{ik} [I_n]_{kj} && \text{Theorem EMP [199]}\\
&= [A]_{ij} [I_n]_{jj} + \sum_{k=1,k\neq j}^{n} [A]_{ik} [I_n]_{kj}\\
&= [A]_{ij} (1) + \sum_{k=1,k\neq j}^{n} [A]_{ik} (0) && \text{Definition IM [76]}\\
&= [A]_{ij} + \sum_{k=1,k\neq j}^{n} 0\\
&= [A]_{ij}
\end{aligned}$$

So the matrices A and AI_n are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [159]) 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. $(B+C)D = BD + CD$ □

Proof We'll do (1), you do (2). Entry-by-entry,

$$\begin{aligned}
[A(B+C)]_{ij} &= \sum_{k=1}^{n} [A]_{ik} [B+C]_{kj} && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} [A]_{ik} ([B]_{kj} + [C]_{kj}) && \text{Definition MA [160]}\\
&= \sum_{k=1}^{n} [A]_{ik} [B]_{kj} + [A]_{ik} [C]_{kj} \\
&= \sum_{k=1}^{n} [A]_{ik} [B]_{kj} + \sum_{k=1}^{n} [A]_{ik} [C]_{kj} \\
&= [AB]_{ij} + [AC]_{ij} && \text{Theorem EMP [199]}\\
&= [AB + AC]_{ij} && \text{Definition MA [160]}
\end{aligned}$$

So the matrices $A(B+C)$ and $AB+AC$ are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [159]) 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 $\alpha(AB) = (\alpha A)B = A(\alpha B)$. □

Proof These are equalities of matrices. We'll do the first one, the second is similar and will be good practice for you.

$$\begin{aligned}
[\alpha(AB)]_{ij} &= \alpha [AB]_{ij} && \text{Definition SMM [160]}\\
&= \alpha \sum_{k=1}^{n} [A]_{ik} [B]_{kj} && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} \alpha [A]_{ik} [B]_{kj} \\
&= \sum_{k=1}^{n} [\alpha A]_{ik} [B]_{kj} && \text{Definition SMM [160]}\\
&= [(\alpha A)B]_{ij} && \text{Theorem EMP [199]}
\end{aligned}$$

So the matrices $\alpha(AB)$ and $(\alpha A)B$ are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [159]) 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.

$$\begin{aligned}
[A(BD)]_{ij} &= \sum_{k=1}^{n} [A]_{ik} [BD]_{kj} && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} [A]_{ik} \left(\sum_{\ell=1}^{p} [B]_{k\ell} [D]_{\ell j} \right) && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} \sum_{\ell=1}^{p} [A]_{ik} [B]_{k\ell} [D]_{\ell j}
\end{aligned}$$

We can switch the order of the summation since these are finite sums,

$$= \sum_{\ell=1}^{p} \sum_{k=1}^{n} [A]_{ik} [B]_{k\ell} [D]_{\ell j}$$

As $[D]_{\ell j}$ does not depend on the index k, we can factor it out of the inner sum,

$$\begin{aligned}
&= \sum_{\ell=1}^{p} [D]_{\ell j} \left(\sum_{k=1}^{n} [A]_{ik} [B]_{k\ell} \right)\\
&= \sum_{\ell=1}^{p} [D]_{\ell j} [AB]_{i\ell} && \text{Theorem EMP [199]}\\
&= \sum_{\ell=1}^{p} [AB]_{i\ell} [D]_{\ell j}\\
&= [(AB)D]_{ij} && \text{Theorem EMP [199]}
\end{aligned}$$

So the matrices $(AB)D$ and $A(BD)$ are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [159]) we can say they are equal matrices. ■

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

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^t \overline{\mathbf{v}}$$
□

Proof Write $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_m \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix}$. Then

$$\begin{aligned}
\langle \mathbf{u},\, \mathbf{v} \rangle &= \sum_{k=1}^{m} u_k \overline{v_k} && \text{Definition IP [146]} \\
&= \sum_{k=1}^{m} [\mathbf{u}]_{k1} \overline{[\mathbf{v}]_{k1}} \\
&= \sum_{k=1}^{m} \left[\mathbf{u}^t\right]_{1k} \overline{[\mathbf{v}]_{k1}} && \text{Definition TM [163]} \\
&= \sum_{k=1}^{m} \left[\mathbf{u}^t\right]_{1k} [\overline{\mathbf{v}}]_{k1} && \text{Definition CCV [145]} \\
&= \left[\mathbf{u}^t \overline{\mathbf{v}}\right]_{11} && \text{Theorem EMP [199]}
\end{aligned}$$

To finish we just blur the distinction between a 1×1 matrix ($\mathbf{u}^t\overline{\mathbf{v}}$) 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,

$$\begin{aligned}
\left[\overline{AB}\right]_{ij} &= \overline{[AB]_{ij}} && \text{Definition CM [71]} \\
&= \overline{\sum_{k=1}^{n} [A]_{ik} [B]_{kj}} && \text{Theorem EMP [199]} \\
&= \sum_{k=1}^{n} \overline{[A]_{ik} [B]_{kj}} && \text{Theorem CCRA [548]} \\
&= \sum_{k=1}^{n} \overline{[A]_{ik}}\, \overline{[B]_{kj}} && \text{Theorem CCRM [548]} \\
&= \sum_{k=1}^{n} \left[\overline{A}\right]_{ik} \left[\overline{B}\right]_{kj} && \text{Definition CCM [165]} \\
&= \left[\overline{A}\,\overline{B}\right]_{ij} && \text{Theorem EMP [199]}
\end{aligned}$$

So the matrices $\overline{AB}$ and $\overline{A}\,\overline{B}$ are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [159]) we can say they are equal matrices. ■

One more theorem in this style, and its 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 $(AB)^t = B^tA^t$. □

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 operation).

Here we go again, entry-by-entry,

$$
\begin{aligned}
\left[(AB)^t\right]_{ij} &= [AB]_{ji} && \text{Definition TM [163]}\\
&= \sum_{k=1}^{n} [A]_{jk}\,[B]_{ki} && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} [B]_{ki}\,[A]_{jk} && \\
&= \sum_{k=1}^{n} \left[B^t\right]_{ik}\left[A^t\right]_{kj} && \text{Definition TM [163]}\\
&= \left[B^tA^t\right]_{ij} && \text{Theorem EMP [199]}
\end{aligned}
$$

So the matrices $(AB)^t$ and B^tA^t are equal, entry-by-entry, and by the definition of matrix equality (Definition ME [159]) 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 is a bonus.

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.

These theorems, along with Theorem VSPM [161], give you the "rules" for how matrices interact with the various operations we have defined. 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 may make us feel that the definition of matrix multiplication is not so unnatural.

Subsection PSHS
Particular Solutions, Homogeneous Solutions

Having delayed presenting matrix multiplication, we have one theorem we could have stated long ago, but its proof is much easier now that we know how to represent a system of linear equations with matrix multiplication and how to mix matrix multiplication with other operations.

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 $\mathbf{z}$ 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 $\mathbf{y} = \mathbf{z} + \mathbf{w}$ for some vector $\mathbf{w} \in N(A)$. □

Proof We will work with the vector equality representations of the relevant systems of equations, as described by Theorem SLEMM [196].

($\Leftarrow$) Suppose $\mathbf{y} = \mathbf{z} + \mathbf{w}$ and $\mathbf{w} \in N(A)$. Then

$$A\mathbf{y} = A(\mathbf{z} + \mathbf{w}) = A\mathbf{z} + A\mathbf{w} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

demonstrating that $\mathbf{y}$ is a solution.

($\Rightarrow$) Suppose $\mathbf{y}$ is a solution to $\mathcal{LS}(A, b)$. Then

$$A(\mathbf{y} - \mathbf{z}) = A\mathbf{y} - A\mathbf{z} = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

which says that $\mathbf{y} - \mathbf{z} \in \mathcal{N}(A)$. In other words, $\mathbf{y} - \mathbf{z} = \mathbf{w}$ for some vector $\mathbf{w} \in N(A)$. Rewritten, this is $\mathbf{y} = \mathbf{z} + \mathbf{w}$, as desired. ■

After proving Theorem NSMUS [79] 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 NSTNS [78]). For a given vector of constants, $\mathbf{b}$, the system $\mathcal{LS}(A, b)$ could be inconsistent, meaning there are no solutions. But if there is at least one solution ($\mathbf{z}$), then Theorem PSPHS [206] tells us there will be infinitely many solutions. So a system of equations with a singular coefficient matrix *never* has a unique solution.

Example PSNS
Particular solutions, homogeneous solutions, Archetype D
Archetype D [487] is a consistent system of equations with a nontrivial null space. The

write-up for this system begins with three solutions,

$$\mathbf{y}_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} \qquad \mathbf{y}_2 = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{y}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \\ 3 \end{bmatrix}$$

We will choose to have $\mathbf{y}_1$ play the role of $\mathbf{z}$ in the statement of Theorem PSPHS [206], 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 $\mathbf{z}$ 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

$$\mathbf{y}_1 = \mathbf{z} = \mathbf{z} + \mathbf{0}.$$

The vectors $\mathbf{y}_2$ and $\mathbf{y}_3$ will require a bit more effort. Solutions to the homogeneous system are exactly the elements of the null space of the coefficient matrix, which is

$$\mathcal{S}p\left(\left\{\begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix},\, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

Then

$$\mathbf{y}_2 = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ -1 \\ -2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} + \left((-2)\begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}\right) = \mathbf{z} + \mathbf{w}_2$$

where

$$\mathbf{w_2} = \begin{bmatrix} 4 \\ -1 \\ -2 \\ -1 \end{bmatrix} = (-2)\begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + (-1)\begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

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 as a span (or as a check, you could just evaluate the equations in the homogeneous system with $\mathbf{w_2}$).

Again

$$\mathbf{y}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 7 \\ 7 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 1 \end{bmatrix} + \left((-1)\begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}\right) = \mathbf{z} + \mathbf{w}_3$$

where

$$\mathbf{w_3} = \begin{bmatrix} 7 \\ 7 \\ -1 \\ 2 \end{bmatrix} = (-1)\begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}$$

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 as a span (or as a check, you could just evaluate the equations in the homogeneous system with $\mathbf{w_2}$).

Here's another view of this theorem, in the context of this example. Grab two new solutions of the original system of equations, say

$$\mathbf{y}_4 = \begin{bmatrix} 11 \\ 0 \\ -3 \\ -1 \end{bmatrix} \qquad \mathbf{y}_5 = \begin{bmatrix} -4 \\ 2 \\ 4 \\ 2 \end{bmatrix}$$

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 [206]. If we let D denote the coefficient matrix then we can use the following application of Theorem PSPHS [206] as the basis of a formal proof of this assertion,

$$\begin{aligned} D(\mathbf{y}_4 - \mathbf{y}_5) &= D\left((z + \mathbf{w}_4) - (z + \mathbf{w}_5)\right) \\ &= D(\mathbf{w}_4 - \mathbf{w}_5) \\ &= D\mathbf{w}_4 - D\mathbf{w}_5 \\ &= \mathbf{0} - \mathbf{0} = \mathbf{0}. \end{aligned}$$

It would be very instructive to formulate the precise statement of a theorem and fill in the details and justifications of the proof. ⊚

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).

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

3. Rewrite the system of linear equations below using matrices and vectors, along with a matrix-vector product.

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 0 \\ x_1 + 2x_2 + x_3 &= 3 \\ x_1 + 3x_2 + 3x_3 &= 7 \end{aligned}$$

Subsection EXC
Exercises

C20 Contributed by Robert Beezer
Compute the product of the two matrices below, AB. Do this using the definitions of the matrix-vector product (Definition MVP [195]) and the definition of matrix multiplication (Definition MM [197]).

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

Solution [213]

T10 Contributed by Robert Beezer
Suppose that A is a square matrix and there is a vector, $\mathbf{b}$, such that $\mathcal{LS}(A,\,b)$ has a unique solution. Prove that A is nonsingular. Give a direct proof (perhaps appealing to Theorem PSPHS [206]) rather than just negating a sentence from the text discussing a similar situation.
Solution [213]

T40 Contributed by Robert Beezer
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 $\mathcal{N}(AB) \not\subseteq \mathcal{N}(B)$. Solution [213]

TODO: Converse with A nonsingular, or four parter, prove or disprove?

Subsection SOL
Solutions

C20 Exercise [211] Contributed by Robert Beezer
By Definition MM [197],

$$AB = \left[\begin{bmatrix} 2 & 5 \\ -1 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \middle| \begin{bmatrix} 2 & 5 \\ -1 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} \middle| \begin{bmatrix} 2 & 5 \\ -1 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} -3 \\ 2 \end{bmatrix} \middle| \begin{bmatrix} 2 & 5 \\ -1 & 3 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \end{bmatrix} \right]$$

Repeated applications of Definition MVP [195] give

$$= \left[1 \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} \middle| 5 \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} \middle| -3 \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} \middle| 4 \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} + (-3) \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} \right]$$

$$= \begin{bmatrix} 12 & 10 & 4 & -7 \\ 5 & -5 & 9 & -13 \\ -2 & 10 & -10 & 14 \end{bmatrix}$$

T10 Exercise [211] Contributed by Robert Beezer
Since $\mathcal{LS}(A,\, b)$ has at least one solution, we can apply Theorem PSPHS [206]. Because the solution is assumed to be unique, the null space of A must be trivial. Then Theorem NSTNS [78] implies that A is nonsingular.

The converse of this statement is a trivial application of Theorem NSMUS [79]. That said, we could extend our NSMxx series of theorems with an added equivalence for nonsingularity, "Given a single vector of constants, $\mathbf{b}$, the system $\mathcal{LS}(A,\, b)$ has a unique solution."

T40 Exercise [211] Contributed by Robert Beezer
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 (Technique SE [21]). Suppose $\mathbf{x} \in \mathcal{N}(B)$. Then we know that $B\mathbf{x} = \mathbf{0}$ by Definition NSM [73]. Consider

$$\begin{aligned} (AB)\mathbf{x} &= A(B\mathbf{x}) && \text{Theorem MMA [202]} \\ &= A\mathbf{0} && \text{Hypothesis} \\ &= \mathbf{0} && \text{Theorem MMZM [200]} \end{aligned}$$

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 NSTNS [78]. Let A be the square zero matrix, $\mathcal{O}$, of the same size. Then $AB = \mathcal{O}B = \mathcal{O}$ by Theorem MMZM [200] and therefore $\mathcal{N}(AB) = \mathbb{C}^n$, and is *not* a subset of $\mathcal{N}(B) = \{\mathbf{0}\}$.

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 [478] is the system of $m = 3$ linear equations in $n = 3$ variables,

$$\begin{aligned} -7x_1 - 6x_2 - 12x_3 &= -33 \\ 5x_1 + 5x_2 + 7x_3 &= 24 \\ x_1 + 4x_3 &= 5 \end{aligned}$$

By Theorem SLEMM [196] we can represent this system of equations as

$$A\mathbf{x} = \mathbf{b}$$

where

$$A = \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} -33 \\ 24 \\ 5 \end{bmatrix}.$$

We'll pull a rabbit out of our hat and present 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 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,

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ B(A\mathbf{x}) &= B\mathbf{b} \\ (BA)\mathbf{x} &= B\mathbf{b} && \text{Theorem MMA [202]} \\ I_3\mathbf{x} &= B\mathbf{b} \\ \mathbf{x} &= B\mathbf{b} && \text{Theorem MMIM [201]} \end{aligned}$$

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 NSMUS [79] 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 in this case left just the vector of unknowns, $\mathbf{x}$, on the left-side of the equation. This is entirely analogous to how we would solve a single linear equation like $3x = 12$. We would multiply both sides by $\frac{1}{3} = 3^{-1}$, the multiplicative inverse of 3. This works fine for any scalar multiple of x, except for zero, which does not have a multiplicative inverse. For matrices, it is 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, and we write $B = A^{-1}$. △

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 [215] the matrix B is the inverse the coefficient matrix of Archetype B [478]. To see this it only remains to check that $AB = I_3$. What about Archetype A [473]? 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 [473],

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

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 [215], this vector equation would have the unique solution $\mathbf{x} = B\mathbf{b}$.

However, this system is inconsistent. Form the augmented matrix $[\,A\,|\,\mathbf{b}]$ and row-reduce to

$$\begin{bmatrix}\boxed{1} & 0 & 1 & 0\\ 0 & \boxed{1} & -1 & 0\\ 0 & 0 & 0 & \boxed{1}\end{bmatrix}$$

which allows to recognize the inconsistency by Theorem RCLS [56].

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.

Its possible this example is less than satisfying. Just where did that particular choice of the vector $\mathbf{b}$ come from anyway? Turns out its not too mysterious. We wanted an inconsistent system, so Theorem RCS [168] suggested choosing a vector *outside* of the range of A (see Example RAA [179] for full disclosure). ⊚

Lets look at one more matrix inverse before we embark on a more systematic study.

Example MIAK
Matrix Inverse, Archetype K

Consider the matrix defined as Archetype K [520],

$$K = \begin{bmatrix}10 & 18 & 24 & 24 & -12\\ 12 & -2 & -6 & 0 & -18\\ -30 & -21 & -23 & -30 & 39\\ 27 & 30 & 36 & 37 & -30\\ 18 & 24 & 30 & 30 & -20\end{bmatrix}.$$

And the matrix

$$L = \begin{bmatrix}1 & -\left(\frac{9}{4}\right) & -\left(\frac{3}{2}\right) & 3 & -6\\ \frac{21}{2} & \frac{43}{4} & \frac{21}{2} & 9 & -9\\ -15 & -\left(\frac{21}{2}\right) & -11 & -15 & \frac{39}{2}\\ 9 & \frac{15}{4} & \frac{9}{2} & 10 & -15\\ \frac{9}{2} & \frac{3}{4} & \frac{3}{2} & 6 & -\left(\frac{19}{2}\right)\end{bmatrix}.$$

Then

$$KL = \begin{bmatrix}10 & 18 & 24 & 24 & -12\\ 12 & -2 & -6 & 0 & -18\\ -30 & -21 & -23 & -30 & 39\\ 27 & 30 & 36 & 37 & -30\\ 18 & 24 & 30 & 30 & -20\end{bmatrix}\begin{bmatrix}1 & -\left(\frac{9}{4}\right) & -\left(\frac{3}{2}\right) & 3 & -6\\ \frac{21}{2} & \frac{43}{4} & \frac{21}{2} & 9 & -9\\ -15 & -\left(\frac{21}{2}\right) & -11 & -15 & \frac{39}{2}\\ 9 & \frac{15}{4} & \frac{9}{2} & 10 & -15\\ \frac{9}{2} & \frac{3}{4} & \frac{3}{2} & 6 & -\left(\frac{19}{2}\right)\end{bmatrix} = \begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 1\end{bmatrix}$$

and

$$LK = \begin{bmatrix} 1 & -\left(\frac{9}{4}\right) & -\left(\frac{3}{2}\right) & 3 & -6 \\ \frac{21}{2} & \frac{43}{4} & \frac{21}{2} & 9 & -9 \\ -15 & -\left(\frac{21}{2}\right) & -11 & -15 & \frac{39}{2} \\ 9 & \frac{15}{4} & \frac{9}{2} & 10 & -15 \\ \frac{9}{2} & \frac{3}{4} & \frac{3}{2} & 6 & -\left(\frac{19}{2}\right) \end{bmatrix} \begin{bmatrix} 10 & 18 & 24 & 24 & -12 \\ 12 & -2 & -6 & 0 & -18 \\ -30 & -21 & -23 & -30 & 39 \\ 27 & 30 & 36 & 37 & -30 \\ 18 & 24 & 30 & 30 & -20 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

so by Definition MI [216], we can say that K is invertible and write $L = K^{-1}$. ⊚

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 MINSM [227] we will have some theorems that allow us to more quickly and easily determine when a matrix is invertible.

Subsection CIM
Computing the Inverse of a Matrix

We will have occasion in this subsection (and later) to reference the following frequently used vectors, so we will make a useful definition now.

Definition SUV
Standard Unit Vectors
Let $\mathbf{e}_i \in \mathbb{C}^m$ denote the column vector that is column i of the $m \times m$ identity matrix I_m. Then the set

$$\{\mathbf{e}_1,\, \mathbf{e}_2,\, \mathbf{e}_3,\, \ldots,\, \mathbf{e}_m\} = \{\mathbf{e}_i \mid 1 \leq i \leq m\}$$

is the set of **standard unit vectors** in $\mathbb{C}^m$. △

Notice that $\mathbf{e}_i$ is a column vector full of zeros, with a lone 1 in the i-th position. We will make reference to these vectors often.

We've seen that the matrices from Archetype B [478] and Archetype K [520] 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, we have

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

□

Proof ($\Leftarrow$) If $ad - bc \neq 0$ then the displayed formula is legitimate (we are not dividing by zero), and it is a simple matter to actually check that $A^{-1}A = AA^{-1} = I_2$.

($\Rightarrow$) Assume that A is invertible, and proceed with a proof by contradiction, by assuming also that $ad - bc = 0$. This means that $ad = bc$. Let

$$B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

be a putative inverse of A. This means that

$$I_2 = AB = \begin{bmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{bmatrix}$$

For the matrix on the right, multiply the top row by c and the bottom row by a. Since we are assuming that $ad = bc$, massage the bottom row by replacing ad by bc in two places. The result is that the two rows of the matrix are identical. Suppose we did the same to I_2, multiply the top row by c and the bottom row by a, and then arrived arrived at equal rows? Given the form of I_2 there is only one way this could happen: $a = 0$ and $c = 0$.

With this information, the product AB simplifies to

$$AB = \begin{bmatrix} bg & bh \\ dg & dh \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

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,.... Note how the above proof never divides, but always multiplies, and how zero/nonzero considerations are handled. Pay attention to the expression $ad - bc$, we will see it again in a while.

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. 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 CMIAK
Computing a Matrix Inverse, Archetype K

Consider the matrix defined as Archetype K [520],

$$A = \begin{bmatrix} 10 & 18 & 24 & 24 & -12 \\ 12 & -2 & -6 & 0 & -18 \\ -30 & -21 & -23 & -30 & 39 \\ 27 & 30 & 36 & 37 & -30 \\ 18 & 24 & 30 & 30 & -20 \end{bmatrix}.$$

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 [197],

$$\begin{aligned} AB &= I_5 \\ A[\mathbf{B}_1|\mathbf{B}_2|\mathbf{B}_3|\mathbf{B}_4|\mathbf{B}_5] &= [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\mathbf{e}_4|\mathbf{e}_5] \\ [A\mathbf{B}_1|A\mathbf{B}_2|A\mathbf{B}_3|A\mathbf{B}_4|A\mathbf{B}_5] &= [\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\mathbf{e}_4|\mathbf{e}_5]. \end{aligned}$$

Equating the matrices column-by-column we have

$$A\mathbf{B}_1 = \mathbf{e}_1 \qquad A\mathbf{B}_2 = \mathbf{e}_2 \qquad A\mathbf{B}_3 = \mathbf{e}_3 \qquad A\mathbf{B}_4 = \mathbf{e}_4 \qquad A\mathbf{B}_5 = \mathbf{e}_5.$$

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 these systems has the same coefficient matrix. We'll now solve each system in turn,

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A,\,\mathbf{e}_1)$,

$$\begin{bmatrix} 10 & 18 & 24 & 24 & -12 & 1 \\ 12 & -2 & -6 & 0 & -18 & 0 \\ -30 & -21 & -23 & -30 & 39 & 0 \\ 27 & 30 & 36 & 37 & -30 & 0 \\ 18 & 24 & 30 & 30 & -20 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & 1 \\ 0 & \boxed{1} & 0 & 0 & 0 & \frac{21}{2} \\ 0 & 0 & \boxed{1} & 0 & 0 & -15 \\ 0 & 0 & 0 & \boxed{1} & 0 & 9 \\ 0 & 0 & 0 & 0 & \boxed{1} & \frac{9}{2} \end{bmatrix} \rightarrow \mathbf{B}_1 = \begin{bmatrix} 1 \\ \frac{21}{2} \\ -15 \\ 9 \\ \frac{9}{2} \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A,\,\mathbf{e}_2)$,

$$\begin{bmatrix} 10 & 18 & 24 & 24 & -12 & 0 \\ 12 & -2 & -6 & 0 & -18 & 1 \\ -30 & -21 & -23 & -30 & 39 & 0 \\ 27 & 30 & 36 & 37 & -30 & 0 \\ 18 & 24 & 30 & 30 & -20 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & -\frac{9}{4} \\ 0 & \boxed{1} & 0 & 0 & 0 & \frac{43}{4} \\ 0 & 0 & \boxed{1} & 0 & 0 & -\frac{21}{2} \\ 0 & 0 & 0 & \boxed{1} & 0 & \frac{15}{4} \\ 0 & 0 & 0 & 0 & \boxed{1} & \frac{3}{4} \end{bmatrix} \rightarrow \mathbf{B}_2 = \begin{bmatrix} -\frac{9}{4} \\ \frac{43}{4} \\ -\frac{21}{2} \\ \frac{15}{4} \\ \frac{3}{4} \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A,\,\mathbf{e}_3)$,

$$\begin{bmatrix} 10 & 18 & 24 & 24 & -12 & 0 \\ 12 & -2 & -6 & 0 & -18 & 0 \\ -30 & -21 & -23 & -30 & 39 & 1 \\ 27 & 30 & 36 & 37 & -30 & 0 \\ 18 & 24 & 30 & 30 & -20 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & -\frac{3}{2} \\ 0 & \boxed{1} & 0 & 0 & 0 & \frac{21}{2} \\ 0 & 0 & \boxed{1} & 0 & 0 & -11 \\ 0 & 0 & 0 & \boxed{1} & 0 & \frac{9}{2} \\ 0 & 0 & 0 & 0 & \boxed{1} & \frac{3}{2} \end{bmatrix} \rightarrow \mathbf{B}_3 = \begin{bmatrix} -\frac{3}{2} \\ \frac{21}{2} \\ -11 \\ \frac{9}{2} \\ \frac{3}{2} \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A,\,\mathbf{e}_4)$,

$$\begin{bmatrix} 10 & 18 & 24 & 24 & -12 & 0 \\ 12 & -2 & -6 & 0 & -18 & 0 \\ -30 & -21 & -23 & -30 & 39 & 0 \\ 27 & 30 & 36 & 37 & -30 & 1 \\ 18 & 24 & 30 & 30 & -20 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & 3 \\ 0 & \boxed{1} & 0 & 0 & 0 & 9 \\ 0 & 0 & \boxed{1} & 0 & 0 & -15 \\ 0 & 0 & 0 & \boxed{1} & 0 & 10 \\ 0 & 0 & 0 & 0 & \boxed{1} & 6 \end{bmatrix} \rightarrow \mathbf{B}_4 = \begin{bmatrix} 3 \\ 9 \\ -15 \\ 10 \\ 6 \end{bmatrix}$$

Row-reduce the augmented matrix of the linear system $\mathcal{LS}(A,\,\mathbf{e}_5)$,

$$\begin{bmatrix} 10 & 18 & 24 & 24 & -12 & 0\\ 12 & -2 & -6 & 0 & -18 & 0\\ -30 & -21 & -23 & -30 & 39 & 0\\ 27 & 30 & 36 & 37 & -30 & 0\\ 18 & 24 & 30 & 30 & -20 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & -6\\ 0 & \boxed{1} & 0 & 0 & 0 & -9\\ 0 & 0 & \boxed{1} & 0 & 0 & \frac{39}{2}\\ 0 & 0 & 0 & \boxed{1} & 0 & -15\\ 0 & 0 & 0 & 0 & \boxed{1} & -\frac{19}{2} \end{bmatrix} \rightarrow \mathbf{B}_5 = \begin{bmatrix} -6\\ -9\\ \frac{39}{2}\\ -15\\ -\frac{19}{2} \end{bmatrix}$$

We can now collect our 5 solution vectors into the matrix B,

$$\begin{aligned} B =&[\mathbf{B}_1|\mathbf{B}_2|\mathbf{B}_3|\mathbf{B}_4|\mathbf{B}_5]\\ =&\left[\begin{bmatrix} 1\\ \frac{21}{2}\\ -15\\ 9\\ \frac{9}{2} \end{bmatrix}\middle|\begin{bmatrix} -\frac{9}{4}\\ \frac{43}{4}\\ -\frac{21}{2}\\ \frac{15}{4}\\ \frac{3}{4} \end{bmatrix}\middle|\begin{bmatrix} -\frac{3}{2}\\ \frac{21}{2}\\ -11\\ \frac{9}{2}\\ \frac{3}{2} \end{bmatrix}\middle|\begin{bmatrix} 3\\ 9\\ -15\\ 10\\ 6 \end{bmatrix}\middle|\begin{bmatrix} -6\\ -9\\ \frac{39}{2}\\ -15\\ -\frac{19}{2} \end{bmatrix}\right] = \begin{bmatrix} 1 & -\left(\frac{9}{4}\right) & -\left(\frac{3}{2}\right) & 3 & -6\\ \frac{21}{2} & \frac{43}{4} & \frac{21}{2} & 9 & -9\\ -15 & -\left(\frac{21}{2}\right) & -11 & -15 & \frac{39}{2}\\ 9 & \frac{15}{4} & \frac{9}{2} & 10 & -15\\ \frac{9}{2} & \frac{3}{4} & \frac{3}{2} & 6 & -\left(\frac{19}{2}\right) \end{bmatrix} \end{aligned}$$

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 CINSM
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 B be the matrix formed from the final n columns of N. Then $AB = I_n$. □

Proof A is nonsingular, so by Theorem NSRRI [76] 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 it 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 system into reduced row-echelon form. Furthermore, the unique solution to each of these systems appears in column $n+1$ of the row-reduced augmented matrix and is equal 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,

$$\begin{aligned} AB =&A[\mathbf{N}_{n+1}|\mathbf{N}_{n+2}|\mathbf{N}_{n+3}|\ldots|\mathbf{N}_{n+n}]\\ =&[A\mathbf{N}_{n+1}|A\mathbf{N}_{n+2}|A\mathbf{N}_{n+3}|\ldots|A\mathbf{N}_{n+n}]\\ =&[\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\ldots|\mathbf{e}_n]\\ =&I_n \end{aligned}$$

as desired. ■

Does this theorem remind you of any others we've seen lately? (Hint: Theorem RNS [177].) We have to be just a bit careful here. This theorem only guarantees that $AB = I_n$, while the definition requires that $BA = I_n$ also. However, we'll soon see that this is *always* the case, in Theorem OSIS [228], so the title of this theorem is not inaccurate.

We'll finish by computing the inverse for the coefficient matrix of Archetype B [478], the one we just pulled from a hat in Example SABMI [215]. There are more examples in the Archetypes (Chapter A [469]) 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 [216]?).

Example CMIAB
Computing a Matrix Inverse, Archetype B
Archetype B [478] has a coefficient matrix given as

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

Exercising Theorem CINSM [221] 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 an inverse, many computational devices have a built-in procedure.

Computation Note MI.MMA
Matrix Inverses (Mathematica)
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 NSI [228]). ⊕

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 [78], 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. Then, repeated use of Definition MI [216] and Theorem MMIM [201] plus one application of Theorem MMA [202] gives

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

and we conclude that B and C cannot be different. So any matrix that acts like the inverse, must be *the* inverse. ■

When most of 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 $(AB)^{-1} = B^{-1}A^{-1}$ and AB is an invertible matrix. □

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), also looking for a date. Lets see if $B^{-1}A^{-1}$ is a good match for AB,

$$\begin{aligned}(B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n\\ (AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.\end{aligned}$$

So the matrix $B^{-1}A^{-1}$ has met all of the requirements to be AB's inverse (date) and we can write $(AB)^{-1} = B^{-1}A^{-1}$. ■

Theorem MIMI
Matrix Inverse of a Matrix Inverse
Suppose A is an invertible matrix. Then $(A^{-1})^{-1} = A$ and A^{-1} is invertible. □

Proof As with the proof of Theorem SS [223], we see if A is a suitable inverse for A^{-1}. Apply Definition MI [216] to see that

$$\begin{aligned} AA^{-1} &= I_n \\ A^{-1}A &= I_n \end{aligned}$$

The matrix A has met all the requirements to be the inverse of A^{-1}, so we can write $(A^{-1})^{-1} = A$. ■

Theorem MIT
Matrix Inverse of a Transpose
Suppose A is an invertible matrix. Then $(A^t)^{-1} = (A^{-1})^t$ and A^t is invertible. □

Proof As with the proof of Theorem SS [223], we see if $(A^{-1})^t$ is a suitable inverse for A^t. Apply Theorem MMT [205] to see that

$$\begin{aligned} (A^{-1})^t A^t &= (AA^{-1})^t = I_n^t = I_n \\ A^t (A^{-1})^t &= (A^{-1}A)^t = I_n^t = I_n \end{aligned}$$

The matrix $(A^{-1})^t$ has met all the requirements to be the inverse of A^t, so 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 [223], we see if $\frac{1}{\alpha}A^{-1}$ is a suitable inverse for αA. Apply Theorem MMSMM [202] to see that

$$\begin{aligned} \left(\frac{1}{\alpha}A^{-1}\right)(\alpha A) &= \left(\frac{1}{\alpha}\alpha\right)\left(AA^{-1}\right) = 1I_n = I_n \\ (\alpha A)\left(\frac{1}{\alpha}A^{-1}\right) &= \left(\alpha\frac{1}{\alpha}\right)\left(A^{-1}A\right) = 1I_n = I_n \end{aligned}$$

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?

Subsection READ
Reading Questions

1. Compute the inverse of the matrix below.

$$\begin{bmatrix} 4 & 10 \\ 2 & 6 \end{bmatrix}$$

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 has the title it does. (Do not just state the theorem, explain the choice of the title making reference to the theorem itself)

Section MINSM
Matrix Inverses and NonSingular Matrices

We saw in Theorem CINSM [221] 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 [216] showed us that the coefficient matrix from Archetype A [473] 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 NSMI
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.

Theorem PWSMS
Product With a Singular Matrix is Singular
Suppose that A or B are matrices of size n, and one, or both, is singular. Then their product, AB, is singular. □

Proof We will use the vector equation representation of the relevant systems of equations throughout the proof (Theorem SLEMM [196]). We'll do the proof in two cases, and its interesting to notice how we break down the cases.

Case 1. Suppose B is singular. Then there is a nontrivial vector $\mathbf{z}$ so that $B\mathbf{z} = \mathbf{0}$. Then

$$(AB)\mathbf{z} = A(B\mathbf{z}) = A\mathbf{0} = \mathbf{0}$$

so we can conclude that AB is singular.

Case 2. Suppose B is nonsingular and A is singular. This is probably not the second case you were expecting. Why not just state the second case as "A is singular"? The best answer is that the proof is easier with the more restrictive assumption that A is singular *and* B is nonsingular. But before we see why, convince yourself that the two cases, as stated, will cover all the possibilities allowed by our hypothesis.

Since A is singular, there is a nontrivial vector $\mathbf{y}$ so that $A\mathbf{y} = \mathbf{0}$. Now consider the linear system $\mathcal{LS}(B, \mathbf{y})$. Since B is nonsingular, the system has a unique solution, which we will call $\mathbf{w}$. We claim $\mathbf{w}$ is not the zero vector. If $\mathbf{w} = \mathbf{0}$, then

$$\mathbf{y} = B\mathbf{w} = B\mathbf{0} = \mathbf{0}$$

contrary to $\mathbf{y}$ being nontrivial. So $\mathbf{w} \neq \mathbf{0}$. The pieces are in place, so here we go,

$$(AB)\mathbf{w} = A(B\mathbf{w}) = A\mathbf{y} = \mathbf{0}$$

which says, since $\mathbf{w}$ is nontrivial, that AB is singular. ■

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 NSRRI [76]). If B is singular, then Theorem PWSMS [227] would imply that I_n is singular, a contradiction. So B must be nonsingular also. Now that we know that B is nonsingular, we can apply Theorem CINSM [221] to assert the existence of a matrix C so that $BC = I_n$. This application of Theorem CINSM [221] 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

$$C = I_nC = (AB)C = A(BC) = AI_n = A.$$

So it happens that the matrix C we just found is really A in disguise. So we can write

$$I_n = BC = BA$$

which is the desired conclusion. ■

So Theorem OSIS [228] tells us that if A is nonsingular, then the matrix B guaranteed by Theorem CINSM [221] 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 CINSM [221] to find an inverse for A. If A is singular, then the procedure in Theorem CINSM [221] will fail as the first n columns of M will not row-reduce to the identity matrix.

This may feel like we are splitting hairs, but its important that we do not make unfounded assumptions. These observations form the next theorem.

Theorem NSI
NonSingularity is Invertibility
Suppose that A is a square matrix. Then A is nonsingular if and only if A is invertible. □

Proof ($\Leftarrow$) Suppose A is invertible, and consider the homogeneous system represented by $A\mathbf{x} = \mathbf{0}$,

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ A^{-1}A\mathbf{x} &= A^{-1}\mathbf{0} \\ I_n\mathbf{x} &= \mathbf{0} \\ \mathbf{x} &= \mathbf{0} \end{aligned}$$

So A has a trivial null space, which is a fancy way of saying that A is nonsingular.

($\Rightarrow$) Suppose A is nonsingular. By Theorem CINSM [221] we find B so that $AB = I_n$. Then Theorem OSIS [228] tells us that $BA = I_n$. So B is A's inverse, and by construction, A is invertible. ■

So 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. Now we can update our list of equivalences for nonsingular matrices (Theorem NSME3 [180]).

Theorem NSME4
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. The range of A is $\mathbb{C}^n$, $\mathcal{R}(A) = \mathbb{C}^n$.
7. A is invertible. □

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 SNSCM
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 NSMUS [79] we know already that $\mathcal{LS}(A,\,\mathbf{b})$ has a unique solution for every choice of $\mathbf{b}$. We need to show that the expression given is indeed a solution. That's easy, just "plug it in" to the corresponding vector equation representation,

$$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}.$$ ■

Subsection OM
Orthogonal Matrices

Definition OM
Orthogonal Matrices
Suppose that Q is a square matrix of size n such that $\left(\overline{Q}\right)^t Q = I_n$. Then we say Q is **orthogonal**. △

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 OM3
Orthogonal matrix of size 3

$$Q = \begin{bmatrix} \frac{1+i}{\sqrt{5}} & \frac{3+2\,i}{\sqrt{55}} & \frac{2+2i}{\sqrt{22}} \\ \frac{1-i}{\sqrt{5}} & \frac{2+2\,i}{\sqrt{55}} & \frac{-3+i}{\sqrt{22}} \\ \frac{i}{\sqrt{5}} & \frac{3-5\,i}{\sqrt{55}} & -\frac{2}{\sqrt{22}} \end{bmatrix}$$

The computations get a bit tiresome, but if you work your way through $\left(\overline{Q}\right)^t Q$, you *will* arrive at the 3×3 identity matrix I_3. ⊚

Orthogonal matrices do not have to look quite so gruesome. Here's a larger one that is a bit more pleasing.

Example OPM
Orthogonal 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 orthogonal 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 [76]).

An interesting exercise is to build another 5×5 orthogonal matrix, R, using a different rearrangement of the columns of I_5. Then form the product PR. This will be another orthogonal matrix (reference exercise here). If you were to build all $5 \times 4 \times 3 \times 2 \times 1! = 120$ 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 it is closed, associative, has an identity (I_5), and inverses (Theorem OMI [230]). Notice though that the operation in this group is not commutative! ⊚

Orthogonal matrices have easily computed inverses. They also have columns that form orthonormal sets. Here are the theorems that show us that orthogonal matrices are not as strange as they might initially appear.

Theorem OMI
Orthogonal Matrices are Invertible
Suppose that Q is an orthogonal matrix of size n. Then Q is nonsingular, and $Q^{-1} = (\overline{Q})^t$. □

Proof By Definition OM [229], we know that $(\overline{Q})^t Q = I_n$. If $(\overline{Q})^t$ or Q were singular, then this equation, together with Theorem PWSMS [227], would have us conclude that I_n is singular, a contradiction, since I_n row-reduces to the identity matrix (Theorem NSRRI [76]). So Q, and $(\overline{Q})^t$, are both nonsingular.

The equation $(\overline{Q})^tQ = I_n$ gets us halfway to an inverse of Q, and since we now know that $(\overline{Q})^t$ is nonsingular, Theorem OSIS [228] tells us that $Q(\overline{Q})^t = I_n$ also. So Q and $(\overline{Q})^t$ are inverses of each other. ■

Theorem COMOS
Columns of Orthogonal Matrices are Orthonormal Sets
Suppose that A is a square matrix of size n with columns $S = \{\mathbf{A}_1,\,\mathbf{A}_2,\,\mathbf{A}_3,\,\ldots,\,\mathbf{A}_n\}$. Then A is an orthogonal matrix if and only if S is an orthonormal set. □

Proof The proof revolves around recognizing that a typical entry of the product $(\overline{A})^tA$ is an inner product of columns of A. Here are the details to support this claim.

$$
\begin{aligned}
\left[\left(\overline{A}\right)^t A\right]_{ij} &= \sum_{k=1}^{n} \left[(\overline{A})^t\right]_{ik} [A]_{kj} && \text{Theorem EMP [199]}\\
&= \sum_{k=1}^{n} \left[\overline{A}\right]_{ki} [A]_{kj} && \text{Definition TM [163]}\\
&= \sum_{k=1}^{n} \overline{[A]_{ki}}\, [A]_{kj} && \text{Definition CCM [165]}\\
&= \sum_{k=1}^{n} [A]_{kj}\, \overline{[A]_{ki}} && \text{Commutativity}\\
&= \langle \mathbf{A}_j,\,\mathbf{A}_i\rangle && \text{Definition IP [146]}
\end{aligned}
$$

$S = \{\mathbf{A}_1,\,\mathbf{A}_2,\,\mathbf{A}_3,\,\ldots,\,\mathbf{A}_n\}$ is an orthonormal set if and only if $\langle \mathbf{A}_j,\,\mathbf{A}_i\rangle = 0$ when $i \neq j$ and $\langle \mathbf{A}_j,\,\mathbf{A}_i\rangle = 1$ when $i = j$ by Definition ONS [156]. However, the above expression for an entry of the matrix product $\left(\overline{A}\right)^t A$ as an inner product means this latter condition is precisely the statement that $\left(\overline{A}\right)^t A = I_n$. ■

Example OSMC
Orthonormal Set from Matrix Columns
The matrix

$$
Q = \begin{bmatrix}
\frac{1+i}{\sqrt{5}} & \frac{3+2\,i}{\sqrt{55}} & \frac{2+2i}{\sqrt{22}}\\
\frac{1-i}{\sqrt{5}} & \frac{2+2\,i}{\sqrt{55}} & \frac{-3+i}{\sqrt{22}}\\
\frac{i}{\sqrt{5}} & \frac{3-5\,i}{\sqrt{55}} & -\frac{2}{\sqrt{22}}
\end{bmatrix}
$$

from Example OM3 [230] is an orthogonal matrix. By Theorem COMOS [231] its columns

$$
\left\{
\begin{bmatrix}\frac{1+i}{\sqrt{5}}\\ \frac{3+2\,i}{\sqrt{55}}\\ \frac{2+2i}{\sqrt{22}}\end{bmatrix},
\begin{bmatrix}\frac{1-i}{\sqrt{5}}\\ \frac{2+2\,i}{\sqrt{55}}\\ \frac{-3+i}{\sqrt{22}}\end{bmatrix}
\begin{bmatrix}\frac{i}{\sqrt{5}}\\ \frac{3-5\,i}{\sqrt{55}}\\ -\frac{2}{\sqrt{22}}\end{bmatrix}
\right\}
$$

form an orthonormal set. You might find checking the six inner products of pairs of these vectors easier than doing the matrix product $\left(\overline{Q}\right)^t Q$. ⊚

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 OMPIP
Orthogonal Matrices Preserve Inner Products
Suppose that Q is an orthogonal matrix of size n and $\mathbf{u}$ and $\mathbf{v}$ are two vectors from $\mathbb{C}^n$. Then

$$\langle Q\mathbf{u},\, Q\mathbf{v}\rangle = \langle \mathbf{u},\, \mathbf{v}\rangle \qquad\qquad \|Q\mathbf{v}\| = \|\mathbf{v}\|. \qquad \square$$

Proof We will be a bit fast and loose with the interplay of conjugates and transposes here, but you should be able to supply some of the missing theorems.

$$\begin{aligned}
\langle Q\mathbf{u},\, Q\mathbf{v}\rangle &= (Q\mathbf{u})^t\overline{Q\mathbf{v}} && \text{Definition IP [146]}\\
&= \mathbf{u}^tQ^t\overline{Q\mathbf{v}} && \text{Theorem MMT [205]}\\
&= \mathbf{u}^tQ^t\overline{Q}\overline{\mathbf{v}} && \text{Theorem MMCC [204]}\\
&= \mathbf{u}^t\overline{\overline{Q}}^t\overline{Q}\overline{\mathbf{v}} && \text{Conjugation twice}\\
&= \mathbf{u}^t\overline{\overline{Q}^t}\overline{Q}\overline{\mathbf{v}} && \\
&= \mathbf{u}^t\overline{\overline{Q}^tQ}\overline{\mathbf{v}} && \text{Theorem MMCC [204]}\\
&= \mathbf{u}^t\overline{I_n}\overline{\mathbf{v}} && \text{Definition OM [229]}\\
&= \mathbf{u}^tI_n\overline{\mathbf{v}} && I_n \text{ has real entries}\\
&= \mathbf{u}^t\overline{\mathbf{v}} && \text{Theorem MMIM [201]}\\
&= \langle \mathbf{u},\, \mathbf{v}\rangle && \text{Definition IP [146]}
\end{aligned}$$

The second conclusion is just a specialization.

$$\begin{aligned}
\|Q\mathbf{v}\|^2 &= \langle Q\mathbf{v},\, Q\mathbf{v}\rangle && \text{Theorem IPN [150]}\\
&= \langle \mathbf{v},\, \mathbf{v}\rangle && \text{Previous conclusion}\\
&= \|\mathbf{v}\|^2 && \text{Theorem IPN [150]}
\end{aligned}$$

Now take a square root on both sides to get the result. ■

Definition A
Adjoint
If A is a square matrix, then its **adjoint** is $A^H = \left(\overline{A}\right)^t$. △

Sometimes a matrix is equal to its adjoint. A simple example would be any symmetric matrix with real entries.

Definition HM
Hermitian Matrix
The square matrix A is **Hermitian** (or **self-adjoint**) if $A = \left(\overline{A}\right)^t$ △

Subsection READ
Reading Questions

1. Show how to use the inverse of a matrix to solve the system of equations below.

$$\begin{aligned} 4x_1 + 10x_2 &= 12 \\ 2x_1 + 6x_2 &= 4 \end{aligned}$$

2. In the previous reading questions you were asked to find the inverse of a 3×3 matrix. Explain your answer to that question in light of a theorem in this section (quote the theorem's acronym).

3. A rare freebie. Write `%#@!` as your solution for full credit.

Subsection EXC
Exercises

C40 Contributed by Robert Beezer
Solve the system of equations below using the inverse of a matrix.

$$\begin{aligned}
x_1 + x_2 + 3x_3 + x_4 &= 5 \\
-2x_1 - x_2 - 4x_3 - x_4 &= -7 \\
x_1 + 4x_2 + 10x_3 + 2x_4 &= 9 \\
-2x_1 - 4x_3 + 5x_4 &= 9
\end{aligned}$$

Solution [237]

TODO: Product of orthogonal matrices is orthogonal (solution via product)
TODO: Hermitian matrices have real entries on the diagonal

Subsection SOL
Solutions

C40 Exercise [235] Contributed by Robert Beezer
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 CINSM [221]. Then Theorem SNSCM [229] says the unique solution is

$$A^{-1}\mathbf{b} = \begin{bmatrix} 38 & 18 & -5 & -2 \\ 96 & 47 & -12 & -5 \\ -39 & -19 & 5 & 2 \\ -16 & -8 & 2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ -7 \\ 9 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \\ 3 \end{bmatrix}$$

VS: Vector Spaces

Section VS
Vector Spaces

We now have a computational toolkit in place and so we can now 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 [379]). Here we present an axiomatic definition of vector spaces, which will lead to an extra increment of abstraction. 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 now.

Subsection VS
Vector Spaces

Here is one of our two most important definitions.

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** if the following ten requirements (better known as "axioms") are met. Let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in V$ and α, $\beta \in \mathbb{C}$.

1. $\mathbf{u} + \mathbf{v} \in V$ (Additive closure)

2. $\alpha\mathbf{u} \in V$ (Scalar closure)

3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (Commutativity)

4. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (Associativity of vector addition)

5. There is a vector, $\mathbf{0} \in V$, called the **zero vector**, such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$ for all $\mathbf{u} \in V$. (Additive identity)

6. For each vector $\mathbf{u} \in V$, there exists a vector $-\mathbf{u} \in V$ so that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$. (Additive inverses)

7. $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$ (Associativity of scalar multiplication)

8. $\alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u}+\alpha\mathbf{v}$ (Distributivity across vector addition)

9. $(\alpha+\beta)\mathbf{u} = \alpha\mathbf{u}+\beta\mathbf{u}$ (Distributivity across addition)

10. $1\mathbf{u}=\mathbf{u}$ (Scalar multiplication with 1)

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. △

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 [241].

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. Here, the use is slightly different. The ten requirements in Definition VS [239] are the most basic properties of the objects relevant to a study of linear algebra, and all of our theorems will be built upon them, as we will begin to see in Subsection VS.VSP [246]. So we will refer to "the vector space axioms." 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 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 axioms are fulfilled (see Example CVS [244]).

A vector space is composed of three objects, a set and two operations. However, 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. In any case, lets look at some concrete examples.

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 axioms and to convince you that the multitude of examples justifies (at least initially) making such a broad definition as Definition VS [239]. Some 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 VSCM
The vector space $\mathbb{C}^m$
Set: $\mathbb{C}^m$, all column vectors of size m, Definition VSCM [89].
Vector Addition: The "usual" addition, given in Definition CVA [91].
Scalar Multiplication: The "usual" scalar multiplication, given in Definition CVSM [92].

Does this set with these operations fulfill the ten axioms? Yes. And all we need to do is quote Theorem VSPCM [94]. That was easy. ⊚

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 [241].
Vector Addition: The "usual" addition, given in Definition MA [160].
Scalar Multiplication: The "usual" scalar multiplication, given in Definition SMM [160].

Does this set with these operations fulfill the ten axioms? Yes. And all we need to do is quote Theorem VSPM [161]. Another easy one. ⊚

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. 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 [89], Chapter M [159]), and both objects, along with their operations, have certain properties in common, as you may have noticed in comparing Theorem VSPCM [94] with Theorem VSPM [161]. Indeed, it is these two theorems that *motivate* us to formulate the abstract definition of a vector space, Definition VS [239]. Now, should we prove some general theorems about vector spaces (as we will shortly in Subsection VS.VSP [246]), 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 polynomials, P_n
Set: P_n, the set of all polynomials of degree n or less in the variable x with coefficients from $\mathbb{C}$.
Vector Addition:

$$\begin{aligned}(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) + (b_0 + b_1x + b_2x^2 + \cdots + b_nx^n) =\\ (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n\end{aligned}$$

Scalar Multiplication:

$$\alpha(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = (\alpha a_0) + (\alpha a_1)x + (\alpha a_2)x^2 + \cdots + (\alpha a_n)x^n$$

This set, with these operations, will fulfill the ten axioms, though we will not work all the details here. However, we will make a few comments and prove one of the axioms. First, the zero vector is what you might expect, and you can check that it has the required property.

$$\mathbf{0} = 0 + 0x + 0x^2 + \cdots + 0x^n$$

The additive inverse is also no surprise, though consider how we have chosen to write it.

$$-\left(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n\right) = (-a_0) + (-a_1)x + (-a_2)x^2 + \cdots + (-a_n)x^n$$

Now lets prove the associativity of vector addition. 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.

$$\begin{aligned}
&\mathbf{u} + (\mathbf{v} + \mathbf{w})\\
&= (a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) + \left((b_0 + b_1x + b_2x^2 + \cdots + b_nx^n) + (c_0 + c_1x + c_2x^2 + \cdots + c_nx^n)\right)\\
&= (a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) + ((b_0 + c_0) + (b_1 + c_1)x + (b_2 + c_2)x^2 + \cdots + (b_n + c_n)x^n)\\
&= (a_0 + (b_0 + c_0)) + (a_1 + (b_1 + c_1))x + (a_2 + (b_2 + c_2))x^2 + \cdots + (a_n + (b_n + c_n))x^n\\
&= ((a_0 + b_0) + c_0) + ((a_1 + b_1) + c_1)x + ((a_2 + b_2) + c_2)x^2 + \cdots + ((a_n + b_n) + c_n)x^n\\
&= ((a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n) + (c_0 + c_1x + c_2x^2 + \cdots + c_nx^n)\\
&= \left((a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n)\right) + (c_0 + c_1x + c_2x^2 + \cdots + c_nx^n)\\
&= (\mathbf{u} + \mathbf{v}) + \mathbf{w}
\end{aligned}$$

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 axioms is similar in style, and tedium. You might try proving the commutativity of vector addition, or one of the distributivity axioms. ⊚

Example VSIS
The vector space of infinite sequences
Set: $\mathbb{C}^\infty = \{\,(c_0,\,c_1,\,c_2,\,c_3,\,\ldots) \mid c_i \in \mathbb{C}\}$.

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? Additive inverses? Can you prove the associativity of vector addition? ⊚

Example VSF
The vector space of functions
Set: $F = \{\, f \mid f : \mathbb{C} \rightarrow \mathbb{C}\}$.
Vector Addition: $f + g$ is the function defined by $(f + g)(x) = f(x) + g(x)$.
Scalar Multiplication: αf is the function defined by $(\alpha f)(x) = \alpha f(x)$.

So this is the set of all functions of one variable that take a complex number 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 study. But since we are allowing our scalars to be complex numbers, we need to expand the domain and range of our functions also. 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 is the function z whose definition is $z(x) = 0$ for every input x.

While vector spaces of functions are very important in mathematics and physics, we will not devote them much more attention.

Here's a unique example.

Example VSS
The singleton vector space
Set: $Z = \{\mathbf{z}\}$.
Vector Addition: $\mathbf{z} + \mathbf{z} = \mathbf{z}$.
Scalar Multiplication: $\alpha\mathbf{z} = \mathbf{z}$.

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*. Our only concern is if this set, along with the definitions of two operations, fulfill the ten axioms. Let's check associativity of vector addition. For all $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in Z$,

$$\begin{aligned}
\mathbf{u} + (\mathbf{v} + \mathbf{w}) &= \mathbf{z} + (\mathbf{z} + \mathbf{z})\\
&= \mathbf{z} + \mathbf{z}\\
&= \mathbf{z}\\
&= \mathbf{z} + \mathbf{z}\\
&= (\mathbf{z} + \mathbf{z}) + \mathbf{z}\\
&= (\mathbf{u} + \mathbf{v}) + \mathbf{w}
\end{aligned}$$

What is the zero vector in this vector space? 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. ⊚

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? Leave your preconceptions behind.

Example CVS
The crazy vector space
Set: $C = \{(x_1,\,x_2) \mid x_1,\,x_2 \in \mathbb{C}\}$.
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 axioms yourself. What is the zero vector? Additive inverses? Can you prove associativity? Here we go.

Additive and scalar closure: The result of each operation is a pair of complex numbers, so these two first axioms are fulfilled.

Commutativity:

$$\begin{aligned}\mathbf{u} + \mathbf{v} &= (x_1,\,x_2) + (y_1,\,y_2) = (x_1 + y_1 + 1,\,x_2 + y_2 + 1)\\ &= (y_1 + x_1 + 1,\,y_2 + x_2 + 1) = (y_1,\,y_2) + (x_1,\,x_2)\\ &= \mathbf{v} + \mathbf{u}\end{aligned}$$

Associativity:

$$\begin{aligned}\mathbf{u} + (\mathbf{v} + \mathbf{w}) &= (x_1,\,x_2) + ((y_1,\,y_2) + (z_1,\,z_2))\\ &= (x_1,\,x_2) + (y_1 + z_1 + 1,\,y_2 + z_2 + 1)\\ &= (x_1 + (y_1 + z_1 + 1) + 1,\,x_2 + (y_2 + z_2 + 1) + 1)\\ &= (x_1 + y_1 + z_1 + 2,\,x_2 + y_2 + z_2 + 2)\\ &= ((x_1 + y_1 + 1) + z_1 + 1,\,(x_2 + y_2 + 1) + z_2 + 1)\\ &= (x_1 + y_1 + 1,\,x_2 + y_2 + 1) + (z_1,\,z_2)\\ &= ((x_1,\,x_2) + (y_1,\,y_2)) + (z_1,\,z_2)\\ &= (\mathbf{u} + \mathbf{v}) + \mathbf{w}\end{aligned}$$

Zero Vector: The zero vector is $\ldots\mathbf{0} = (-1,\,-1)$. 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?

Additive Inverse: For each vector, $\mathbf{u}$, we must locate an 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:

$$\mathbf{u}+(-\mathbf{u}) = (x_1,\,x_2)+(-x_1-2,\,-x_2-2) = (x_1+(-x_1-2)+1,\,-x_2+(x_2-2)+1) = (-1,\,-1) = \mathbf{0}$$

Associativity of Scalar Multiplication:

$$\begin{aligned}
\alpha(\beta\mathbf{u}) &= \alpha(\beta(x_1,\, x_2)) \\
&= \alpha(\beta x_1 + \beta - 1,\, \beta x_2 + \beta - 1)) \\
&= (\alpha(\beta x_1 + \beta - 1) + \alpha - 1,\, \alpha(\beta x_2 + \beta - 1) + \alpha - 1)) \\
&= ((\alpha\beta x_1 + \alpha\beta - \alpha) + \alpha - 1,\, (\alpha\beta x_2 + \alpha\beta - \alpha) + \alpha - 1)) \\
&= (\alpha\beta x_1 + \alpha\beta - 1,\, \alpha\beta x_2 + \alpha\beta - 1)) \\
&= (\alpha\beta)(x_1,\, x_2) \\
&= (\alpha\beta)\mathbf{u}
\end{aligned}$$

Distributivity Across Vector Addition: If you have hung on so far, here's where it gets even wilder. In the next two axioms we mix and mash the two operations.

$$\begin{aligned}
\alpha(\mathbf{u} + \mathbf{v}) &= \alpha\left((x_1,\, x_2) + (y_1,\, y_2)\right) \\
&= \alpha(x_1 + y_1 + 1,\, x_2 + y_2 + 1) \\
&= (\alpha(x_1 + y_1 + 1) + \alpha - 1,\, \alpha(x_2 + y_2 + 1) + \alpha - 1) \\
&= (\alpha x_1 + \alpha y_1 + \alpha + \alpha - 1,\, \alpha x_2 + \alpha y_2 + \alpha + \alpha - 1) \\
&= (\alpha x_1 + \alpha - 1 + \alpha y_1 + \alpha - 1 + 1,\, \alpha x_2 + \alpha - 1 + \alpha y_2 + \alpha - 1 + 1) \\
&= ((\alpha x_1 + \alpha - 1) + (\alpha y_1 + \alpha - 1) + 1,\, (\alpha x_2 + \alpha - 1) + (\alpha y_2 + \alpha - 1) + 1) \\
&= (\alpha x_1 + \alpha - 1,\, \alpha x_2 + \alpha - 1) + (\alpha y_1 + \alpha - 1,\, \alpha y_2 + \alpha - 1) \\
&= \alpha(x_1,\, x_2) + \alpha(y_1,\, y_2) \\
&= \alpha\mathbf{u} + \alpha\mathbf{v}
\end{aligned}$$

Distributivity Across Addition:

$$\begin{aligned}
(\alpha + \beta)\mathbf{u} &= (\alpha + \beta)(x_1,\, x_2) \\
&= ((\alpha + \beta)x_1 + (\alpha + \beta) - 1,\, (\alpha + \beta)x_2 + (\alpha + \beta) - 1) \\
&= (\alpha x_1 + \beta x_1 + \alpha + \beta - 1,\, \alpha x_2 + \beta x_2 + \alpha + \beta - 1) \\
&= (\alpha x_1 + \alpha - 1 + \beta x_1 + \beta - 1 + 1,\, \alpha x_2 + \alpha - 1 + \beta x_2 + \beta - 1 + 1) \\
&= ((\alpha x_1 + \alpha - 1) + (\beta x_1 + \beta - 1) + 1,\, (\alpha x_2 + \alpha - 1) + (\beta x_2 + \beta - 1) + 1) \\
&= (\alpha x_1 + \alpha - 1,\, \alpha x_2 + \alpha - 1) + (\beta x_1 + \beta - 1,\, \beta x_2 + \beta - 1) \\
&= \alpha(x_1,\, x_2) + \beta(x_1,\, x_2) \\
&= \alpha\mathbf{u} + \beta\mathbf{u}
\end{aligned}$$

Multiplication by 1: 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. ⊚

Subsection VSP
Vector Space Properties

Subsection VS.EVS [241] 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 axioms, *and nothing more*. Its 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 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 [244]), 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 [164].)

First we show that there is just one zero vector. Notice that the axioms only require there to be one, and say nothing about there being more. That is because we can use the axioms to learn that there can *never* be more than one. To require that this extra condition be stated in the axioms would make them more complicated than they need 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 [78]). So let $\mathbf{0}_1$ and $\mathbf{0}_2$ be two zero vectors in V. Then

$$\begin{aligned} \mathbf{0}_1 &= \mathbf{0}_1 + \mathbf{0}_2 && \text{Property of the zero vector } \mathbf{0}_2 \\ &= \mathbf{0}_2 && \text{Property of the zero vector } \mathbf{0}_1 \end{aligned}$$

Notice that we have implicitly used the commutativity of vector addition so that we can apply the defining property of a zero vector on either side of the addition. 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. □

Proof To prove uniqueness, a standard technique is to suppose the existence of two objects (Technique U [78]). So let $-\mathbf{u}_1$ and $-\mathbf{u}_2$ be two additive inverses for $\mathbf{u}$. Then

$$\begin{aligned}
-\mathbf{u}_1 &= -\mathbf{u}_1 + \mathbf{0} && \text{Property of the zero vector}\\
&= -\mathbf{u}_1 + (\mathbf{u} + -\mathbf{u}_2) && -\mathbf{u}_2 \text{ is an additive inverse}\\
&= (-\mathbf{u}_1 + \mathbf{u}) + -\mathbf{u}_2 && \text{Associativity of vector addition}\\
&= \mathbf{0} + -\mathbf{u}_2 && -\mathbf{u}_1 \text{ is an additive inverse}\\
&= -\mathbf{u}_2 && \text{Property of the zero vector}
\end{aligned}$$

So the two additive inverses are really the same. ■

As obvious as the next theorem appears, it is not 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

$$\begin{aligned}
0\mathbf{u} &= (0 + 0)\mathbf{u}\\
0\mathbf{u} &= 0\mathbf{u} + 0\mathbf{u} && \text{Distributivity across addition}
\end{aligned}$$

0 is a scalar, $\mathbf{u}$ is a vector, so scalar closure means $0\mathbf{u}$ is again a vector. As such, it has an additive inverse.

$$\begin{aligned}
-(0\mathbf{u}) + 0\mathbf{u} &= -(0\mathbf{u}) + (0\mathbf{u} + 0\mathbf{u})\\
-(0\mathbf{u}) + 0\mathbf{u} &= (-(0\mathbf{u}) + 0\mathbf{u}) + 0\mathbf{u} && \text{Associativity of vector addition}\\
\mathbf{0} &= \mathbf{0} + 0\mathbf{u} && \text{Additive inverses}\\
\mathbf{0} &= 0\mathbf{u} && \text{Property of zero vector}
\end{aligned}$$
■

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

$$\begin{aligned}
\alpha\mathbf{0} &= \alpha(\mathbf{0} + \mathbf{0})\\
\alpha\mathbf{0} &= \alpha\mathbf{0} + \alpha\mathbf{0} && \text{Distributivity across vector addition}
\end{aligned}$$

α is a scalar, $\mathbf{0}$ is a vector, so scalar closure means $\alpha\mathbf{0}$ is again a vector. As such, it has an additive inverse.

$$\begin{aligned}
-(\alpha\mathbf{0}) + \alpha\mathbf{0} &= -(\alpha\mathbf{0}) + (\alpha\mathbf{0} + \alpha\mathbf{0})\\
-(\alpha\mathbf{0}) + \alpha\mathbf{0} &= (-(\alpha\mathbf{0}) + \alpha\mathbf{0}) + \alpha\mathbf{0} && \text{Associativity of vector addition}\\
\mathbf{0} &= \mathbf{0} + \alpha\mathbf{0} && \text{Additive inverses}\\
\mathbf{0} &= \alpha\mathbf{0} && \text{Property of the zero vector}
\end{aligned}$$
■

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 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 $-\mathbf{u}=(-1)\mathbf{u}$. □

Proof

$$
\begin{aligned}
(-1)\mathbf{u}+\mathbf{u} &= (-1)\mathbf{u}+1\mathbf{u} && \text{Scalar multiplication with 1}\\
&= ((-1)+1)\mathbf{u} && \text{Distributivity across addition}\\
&= 0\mathbf{u} &&\\
&= \mathbf{0} && \text{Theorem ZSSM [247]}
\end{aligned}
$$

$(-1)\mathbf{u}$ is a vector in V by scalar closure, and this equation says it acts as the additive inverse of $\mathbf{u}$. Since the additive inverse of a vector is unique (Theorem AIU [246]), $(-1)\mathbf{u}$ must *be* the additive inverse of $\mathbf{u}$. In other words, $-\mathbf{u}=(-1)\mathbf{u}$. ■

Because of this theorem, we will 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**.

Example PCVS
Properties for the Crazy Vector Space
Several of the above theorems have interesting demonstrations when applied to the crazy vector space, C (Example CVS [244]). We are not proving anything new here, or learning anything we did not know already about C. It is just 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 by Theorem ZSSM [247],

$$0\mathbf{u}=0(x_1,\,x_2)=(0x_1+0-1,\,0x_2+0-1)=(-1,-1)=\mathbf{0}$$

By Theorem ZVSM [247],

$$\alpha\mathbf{0}=\alpha(-1,\,-1)=(\alpha(-1)+\alpha-1,\,\alpha(-1)+\alpha-1)=(-\alpha+\alpha-1,\,-\alpha+\alpha-1)=(-1,\,-1)$$

By Theorem AISM [248],

$$(-1)\mathbf{u}=(-1)(x_1,\,x_2)=((-1)x_1+(-1)-1,\,(-1)x_2+(-1)-1)=(-x_1-2,\,-x_2-2)=-\mathbf{u}$$

⊚

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 move from one statement to another.

Theorem SMEZV
Scalar Multiplication Equals the Zero Vector
Suppose that V is a vector space and $\alpha \in \mathbb{C}$. Then 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 satisfied) and we are done. That was easy.

Case 2. Suppose $\alpha \neq 0$.

$$
\begin{aligned}
\alpha\mathbf{u} &= \mathbf{0} && \text{Hypothesis}\\
\frac{1}{\alpha}(\alpha\mathbf{u}) &= \frac{1}{\alpha}\mathbf{0} && \alpha \neq 0\\
\left(\frac{1}{\alpha}\alpha\right)\mathbf{u} &= \frac{1}{\alpha}\mathbf{0} && \text{Associativity of scalar multiplication}\\
1\mathbf{u} &= \mathbf{0} && \text{Theorem ZVSM [247]}\\
\mathbf{u} &= \mathbf{0} && \text{Scalar multiplication with 1}
\end{aligned}
$$

So in this case, the conclusion is true (the second part of the either/or) and we are done since the conclusion was true in each of the cases. ■

The next three theorems give us cancellation properties. The two concerned with scalar multiplication are intimately connected with Theorem SMEZV [249]. All three are implications. So we will prove each once, here and now, and then we can apply them at will in the future, saving several steps in a proof whenever we do.

Theorem VAC
Vector Addition Cancellation
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 $\mathbf{u} = \mathbf{v}$. □

Proof

$$
\begin{aligned}
\mathbf{w} + \mathbf{u} &= \mathbf{w} + \mathbf{v}\\
-\mathbf{w} + (\mathbf{w} + \mathbf{u}) &= -\mathbf{w} + (\mathbf{w} + \mathbf{v}) && \text{Additive inverses exist}\\
(-\mathbf{w} + \mathbf{w}) + \mathbf{u} &= (-\mathbf{w} + \mathbf{w}) + \mathbf{v} && \text{Associativity of vector addition}\\
\mathbf{0} + \mathbf{u} &= \mathbf{0} + \mathbf{v} && \text{Additive inverses}\\
\mathbf{u} &= \mathbf{v} && \text{Property of the zero vector}
\end{aligned}
$$

■

Theorem CSSM
Canceling Scalars in Scalar Multiplication
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 $\mathbf{u} = \mathbf{v}$. □

Proof

$$\begin{aligned}
\alpha\mathbf{u} &= \alpha\mathbf{v} && \text{Hypothesis}\\
\frac{1}{\alpha}(\alpha\mathbf{u}) &= \frac{1}{\alpha}(\alpha\mathbf{v}) && \alpha \neq 0\\
\left(\frac{1}{\alpha}\alpha\right)\mathbf{u} &= \left(\frac{1}{\alpha}\alpha\right)\mathbf{v} && \text{Associativity of scalar multiplication}\\
1\mathbf{u} &= 1\mathbf{v}\\
\mathbf{u} &= \mathbf{v} && \text{Scalar multiplication with 1}
\end{aligned}$$

■

Theorem CVSM
Canceling Vectors in Scalar Multiplication
Suppose V is a vector space, $\mathbf{u} \neq \mathbf{0}$ is a vector in V and α, $\beta \in \mathbb{C}$. If $\alpha\mathbf{u} = \beta\mathbf{u}$, then $\alpha = \beta$. □

Proof

$$\begin{aligned}
\alpha\mathbf{u} &= \beta\mathbf{u} && \text{Hypothesis}\\
-(\alpha\mathbf{u}) + \alpha\mathbf{u} &= -(\alpha\mathbf{u}) + \beta\mathbf{u} && \text{Additive inverses exist}\\
\mathbf{0} &= -(\alpha\mathbf{u}) + \beta\mathbf{u} && \text{Additive inverse}\\
\mathbf{0} &= (-1)(\alpha\mathbf{u}) + \beta\mathbf{u} && \text{Theorem AISM [248]}\\
\mathbf{0} &= ((-1)\alpha)\mathbf{u} + \beta\mathbf{u} && \text{Associativity of scalar multiplication}\\
\mathbf{0} &= ((-\alpha) + \beta)\mathbf{u} && \text{Distributivity across addition}
\end{aligned}$$

Applying Theorem SMEZV [249], along with the hypothesis that $\mathbf{u} \neq \mathbf{0}$ gives

$$\begin{aligned}
-\alpha + \beta &= 0\\
\alpha &= \beta
\end{aligned}$$

■

So with these three theorems in hand, we can return to our practice of "slashing" out parts of an equation, so long as we are careful about not canceling a scalar that might possibly be zero, or canceling a vector in a scalar multiplication that might be the zero vector.

Subsection RD
Recycling Definitions

When we say that V is a vector space, we know we have a set of objects, but we also know we have been provided with two operations. 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 [89] 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. (See Definition LCCV [97] and Definition LC [259], Definition SSCV [117] and Definition SS [260], Definition RLDCV [129] and Definition RLD [271], Definition LICV [129] and Definition LI [271].)

Subsection READ
Reading Questions

1. Comment on how the vector space $\mathbb{C}^m$ went from a theorem (Theorem VSPCM [94]) to an example (Example VSCM [241]).

2. In the crazy vector space, C, (Example CVS [244]) compute the linear combination

$$2(3,\,4) + (-6)(1,\,2).$$

3. Suppose that α is a scalar and $\mathbf{0}$ is the zero vector. Why should we prove anything as obvious as $\alpha\mathbf{0} = \mathbf{0}$ as we did in Theorem ZVSM [247]?

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. △

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 VSCM [241]). 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 $W \subseteq V$, since the objects in W are column vectors of size 3. But is W a vector space? Does it satisfy the ten axioms of Definition VS [239] 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 $\mathbf{x}$ must satisfy $2x_1 - 5x_2 + 7x_3 = 0$ while $\mathbf{y}$ must satisfy $2y_1 - 5y_2 + 7y_3 = 0$. Our first axiom 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,

$$\begin{aligned}
&2(x_1 + y_1) - 5(x_2 + y_2) + 7(x_3 + y_3) \\
&= 2x_1 + 2y_1 - 5x_2 - 5y_2 + 7x_3 + 7y_3 \\
&= (2x_1 - 5x_2 + 7x_3) + (2y_1 - 5y_2 + 7y_3) \\
&= 0 + 0 && \mathbf{x} \in W,\ \mathbf{y} \in W \\
&= 0
\end{aligned}$$

and by this computation we see that $\mathbf{x}+\mathbf{y}\in W$. One axiom down, nine to go.

If α is a scalar and $\mathbf{x}\in W$, is it always true that $\alpha\mathbf{x}\in W$? 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

$$\begin{aligned}&2(\alpha x_1) - 5(\alpha x_2) + 7(\alpha x_3)\\&= \alpha(2x_1 - 5x_2 + 7x_3)\\&= \alpha 0 && \mathbf{x}\in W\\&= 0\end{aligned}$$

and we see that indeed $\alpha\mathbf{x}\in W$. Always.

If W has a zero vector, it will be unique (Theorem ZVU [246]). 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.

With a zero vector, we can now ask about additive inverses. 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

$$\begin{aligned}&2(-x_1) - 5(-x_2) + 7(-x_3)\\&= -(2x_1 - 5x_2 + 7x_3)\\&= -0 \qquad \mathbf{x}\in W\\&= 0\end{aligned}$$

and we now believe that $-\mathbf{x}\in W$.

Is the vector addition in W commutative? 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 axioms are unaffected by the transition to a smaller set of vectors, and so remain true. That was convenient.

So W satisfies all ten axioms, 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 [253] we proceeded through all ten of the vector space axioms before believing that a subset was a subspace. But six of the axioms 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. Whenever $\mathbf{x} \in W$ and $\mathbf{y} \in W$, then $\mathbf{x} + \mathbf{y} \in W$.
3. Whenever $\alpha \in \mathbb{C}$ and $\mathbf{x} \in W$, then $\alpha\mathbf{x} \in W$. □

Proof ($\Rightarrow$) We have the hypothesis that W is a subspace, so by Definition VS [239] 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 axioms, and so exactly meets the second and third conditions. If that was easy, the 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 the first two closure axioms. The commutativity, associativity (twice), distributivity (twice) and multiplication with 1 axioms that hold for V will continue to hold in W since passing to a subset leaves their statements unchanged. Eight down, two to go.

Since W is non-empty, we can choose some vector $\mathbf{z} \in W$. Then by scalar closure (the third part of our hypothesis), we know $(-1)\mathbf{z} \in W$. By Theorem AISM [248] $(-1)\mathbf{z} = -\mathbf{z}$. Now apply the additive inverse axiom for V and then additive closure (the second part of our hypothesis) to 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. So W has a zero vector. One axiom left.

Suppose $\mathbf{x} \in W$. Then by scalar closure (the third part of our hypothesis), we know that $(-1)\mathbf{x} \in W$. By Theorem AISM [248] $(-1)\mathbf{x} = -\mathbf{x}$, 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.

Three conditions, plus being a subset, gets us all ten axioms. Fabulous! ■

This theorem is 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 [253] 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 [242]). Define a subset W as

$$W = \{\, p(x) \mid p \in P_4,\ p(2) = 0\}$$

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,

$$\begin{aligned}(p+q)(2) &= p(2) + q(2) && \text{Addition in } P_4\\ &= 0 + 0 && p \in W,\ q \in W\\ &= 0\end{aligned}$$

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,

$$\begin{aligned}(\alpha p)(2) &= \alpha p(2) && \text{Scalar multiplication in } P_4\\ &= \alpha 0 && p \in W\\ &= 0\end{aligned}$$

so $\alpha p \in W$.

We have shown that W meets the three conditions of Theorem TSS [255] and so qualifies as a subspace of P_4. Notice that by Definition S [253] we now know that W is also a vector space. So all the axioms of a vector space (Definition VS [239]) and the theorems of Section VS [239] apply in full. ⊚

Much of the power of Theorem TSS [255] 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 Section VS [239].

It can be as instructive to consider some subsets that are *not* subspaces. Since Theorem TSS [255] is an equivalence (see Technique E [54]) 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 axioms in Definition VS [239]. 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 the zero vector axiom of Definition VS [239]. 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 Definition VS [239] or Theorem TSS [255], and is therefore not a subspace. ⊚

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 $\mathbb{C}^2$

$$Y = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \middle|\, x_1 \in \mathbb{Z},\, x_2 \in \mathbb{Z} \right\}$$

$\mathbb{Z}$ 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 Definition VS [239] or Theorem TSS [255], and is therefore not a subspace. ⊚

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 [253], V qualifies as a subspace of itself. The set containing just the zero vector $Z = \{\mathbf{0}\}$ is also a subspace as can be seen by applying Theorem TSS [255] or by simple modifications of the techniques hinted at in Example VSS [243]. 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 $\{\mathbf{0}\}$ are each called a **trivial subspace**.△

We can also use Theorem TSS [255] 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 $\mathbb{C}^n$. □

Proof We will examine the three requirements of Theorem TSS [255]. Recall that $\mathcal{N}(A) = \{\mathbf{x} \in \mathbb{C}^n \mid A\mathbf{x} = \mathbf{0}\}$.

First, $\mathbf{0} \in \mathcal{N}(A)$, which can be inferred as a consequence of Theorem HSC [68]. 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.

$$\begin{aligned} A(\mathbf{x}+\mathbf{y}) &= A\mathbf{x} + A\mathbf{y} && \text{Theorem MMDAA [201]} \\ &= \mathbf{0} + \mathbf{0} && \mathbf{x} \in \mathcal{N}(A)\,,\ \mathbf{y} \in \mathcal{N}(A) \\ &= \mathbf{0} && \text{Theorem VSPCM [94]} \end{aligned}$$

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.

$$\begin{aligned} A(\alpha\mathbf{x}) &= \alpha(A\mathbf{x}) && \text{Theorem MMSMM [202]} \\ &= \alpha\mathbf{0} && \mathbf{x} \in \mathcal{N}(A) \\ &= \mathbf{0} && \text{Theorem ZVSM [247]} \end{aligned}$$

So, yes, $\alpha\mathbf{x}$ qualifies for membership in $\mathcal{N}(A)$.

Having met the three conditions in Theorem TSS [255] we can now say that the null space of a matrix is a subspace (and hence a vector space in its own right!). ■

Here is an example where we can exercise Theorem NSMS [258].

Example RSNS
Recasting a subspace as a null space
Consider the subset of $\mathbb{C}^5$ defined as

$$W = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} \middle| \begin{array}{l} 3x_1 + x_2 - 5x_3 + 7x_4 + x_5 = 0, \\ 4x_1 + 6x_2 + 3x_3 - 6x_4 - 5x_5 = 0, \\ -2x_1 + 4x_2 + 7x_4 + x_5 = 0 \end{array} \right\}$$

It is possible to show that W is a subspace of $\mathbb{C}^5$ by checking the three conditions of Theorem TSS [255] 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 [258] 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 [89] and Chapter M [159]. 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 [97] and Definition SSCV [117].

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 α_1, α_2, α_3, $\ldots$, α_n, their **linear combination** is the vector

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n.$$ △

Example LCM
A linear combination of matrices
In the vector space M_{23} of 2×3 matrices, we have the vectors

$$\mathbf{x} = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 0 & 7 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 3 & -1 & 2 \\ 5 & 5 & 1 \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} 4 & 2 & -4 \\ 1 & 1 & 1 \end{bmatrix}$$

and we can form linear combinations such as

$$\begin{aligned}
2\mathbf{x}+4\mathbf{y}+(-1)\mathbf{z} &= 2\begin{bmatrix}1&3&-2\\2&0&7\end{bmatrix}+4\begin{bmatrix}3&-1&2\\5&5&1\end{bmatrix}+(-1)\begin{bmatrix}4&2&-4\\1&1&1\end{bmatrix}\\
&=\begin{bmatrix}2&6&-4\\4&0&14\end{bmatrix}+\begin{bmatrix}12&-4&8\\20&20&4\end{bmatrix}+\begin{bmatrix}-4&-2&4\\-1&-1&-1\end{bmatrix}\\
&=\begin{bmatrix}10&0&8\\23&19&17\end{bmatrix}
\end{aligned}$$

or,

$$\begin{aligned}
4\mathbf{x}-2\mathbf{y}+3\mathbf{z} &= 4\begin{bmatrix}1&3&-2\\2&0&7\end{bmatrix}-2\begin{bmatrix}3&-1&2\\5&5&1\end{bmatrix}+3\begin{bmatrix}4&2&-4\\1&1&1\end{bmatrix}\\
&=\begin{bmatrix}4&12&-8\\8&0&28\end{bmatrix}+\begin{bmatrix}-6&2&-4\\-10&-10&-2\end{bmatrix}+\begin{bmatrix}12&6&-12\\3&3&3\end{bmatrix}\\
&=\begin{bmatrix}10&20&-24\\1&-7&29\end{bmatrix}
\end{aligned}$$

⊚

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**, $\mathcal{S}p(S)$, is the set of all possible linear combinations of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t$. Symbolically,

$$\begin{aligned}
\mathcal{S}p(S) &= \{\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2+\alpha_3\mathbf{u}_3+\cdots+\alpha_t\mathbf{u}_t \mid \alpha_i\in\mathbb{C},\ 1\le i\le t\}\\
&=\left\{\sum_{i=1}^{t}\alpha_i\mathbf{u}_i \,\middle|\, \alpha_i\in\mathbb{C},\ 1\le i\le t\right\}
\end{aligned}$$

△

Theorem SSS
Span of a Set is a Subspace
Suppose V is a vector space. Given a set of vectors $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_t\} \subseteq V$, their span, $\mathcal{S}p(S)$, is a subspace. □

Proof We will verify the three conditions of Theorem TSS [255]. First, use Theorem ZSSM [247] repeatedly and the defining property of the zero vector from Definition VS [239],

$$0\mathbf{u}_1+0\mathbf{u}_2+0\mathbf{u}_3+\cdots+0\mathbf{u}_t+=\mathbf{0}+\mathbf{0}+\mathbf{0}+\ldots+\mathbf{0}=\mathbf{0}$$

so $\mathbf{0}\in\mathcal{S}p(S)$ since we have written $\mathbf{0}$ as a linear combination of the vectors in S.

Second, suppose $\mathbf{x} \in \mathcal{S}p(S)$ and $\mathbf{y} \in \mathcal{S}p(S)$. Can we conclude that $\mathbf{x}+\mathbf{y} \in \mathcal{S}p(S)$? What do we know about $\mathbf{x}$ and $\mathbf{y}$ by virtue of their membership in $\mathcal{S}p(S)$? There must be scalars from $\mathbb{C}$, α_1, α_2, α_3, $\ldots$, α_t and β_1, β_2, β_3, $\ldots$, β_t so that

$$\begin{aligned}\mathbf{x} &= \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_t\mathbf{u}_t\\ \mathbf{y} &= \beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \beta_3\mathbf{u}_3 + \cdots + \beta_t\mathbf{u}_t\end{aligned}$$

Then

$$\begin{aligned}\mathbf{x}+\mathbf{y} &= \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_t\mathbf{u}_t\\ &\quad + \beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \beta_3\mathbf{u}_3 + \cdots + \beta_t\mathbf{u}_t\\ &= \alpha_1\mathbf{u}_1 + \beta_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \beta_2\mathbf{u}_2\\ &\quad + \alpha_3\mathbf{u}_3 + \beta_3\mathbf{u}_3 + \cdots + \alpha_t\mathbf{u}_t + \beta_t\mathbf{u}_t && \text{Associativity, Commutativity}\\ &= (\alpha_1+\beta_1)\mathbf{u}_1 + (\alpha_2+\beta_2)\mathbf{u}_2\\ &\quad + (\alpha_3+\beta_3)\mathbf{u}_3 + \cdots + (\alpha_t+\beta_t)\mathbf{u}_t && \text{Distributivity}\end{aligned}$$

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 we can say that $\mathbf{x}+\mathbf{y} \in \mathcal{S}p(S)$.

Third, suppose $\alpha \in \mathbb{C}$ and $\mathbf{x} \in \mathcal{S}p(S)$. Can we conclude that $\alpha\mathbf{x} \in \mathcal{S}p(S)$? What do we know about $\mathbf{x}$ by virtue of its membership in $\mathcal{S}p(S)$? There must be scalars from $\mathbb{C}$, α_1, α_2, α_3, $\ldots$, α_t so that

$$\mathbf{x} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_t\mathbf{u}_t$$

Then

$$\begin{aligned}\alpha\mathbf{x} &= \alpha\left(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_t\mathbf{u}_t\right)\\ &= \alpha(\alpha_1\mathbf{u}_1) + \alpha(\alpha_2\mathbf{u}_2) + \alpha(\alpha_3\mathbf{u}_3) + \cdots + \alpha(\alpha_t\mathbf{u}_t) && \text{Distributivity}\\ &= (\alpha\alpha_1)\mathbf{u}_1 + (\alpha\alpha_2)\mathbf{u}_2 + (\alpha\alpha_3)\mathbf{u}_3 + \cdots + (\alpha\alpha_t)\mathbf{u}_t && \text{Associativity}\end{aligned}$$

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 we can say that $\alpha\mathbf{x} \in \mathcal{S}p(S)$.

With the three conditions of Theorem TSS [255] met, we can say that $\mathcal{S}p(S)$ is a subspace (and so is a vector space). ■

Example SSP
Span of a set of polynomials

In Example SP4 [256] we proved that

$$W = \{p(x) \mid p \in P_4,\ p(2) = 0\}$$

is a subspace of P_4, the vector space of polynomials of degree at most 4. Since W is a vector space itself, lets construct a span within W. First let

$$S = \left\{x^4 - 16,\, x^2 - x - 2\right\}$$

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 $U = \mathcal{S}p(S)$, then Theorem SSS [260] 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, lets try to gain a better understanding of just what the span construction creates subspaces, in the context of this example. We can quickly build representative elements of U,

$$3(x^4 - 16) + 5(x^2 - x - 2) = 3x^4 + 5x^2 - 5x - 58$$

and

$$(-2)(x^4 - 16) + 16(x^2 - x - 2) = -2x^4 + 16x^2 - 16x$$

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 $\mathbf{y} = x^4+2x^2-5x-14$ 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 $U = \mathcal{S}p(S)$? It must be a linear combination of the vectors in S, $x^4 - 16$ and $x^2 - x - 2$. So lets suppose that $\mathbf{y}$ is such a linear combination,

$$\begin{aligned} x^4 + 2x^2 - 5x - 14 = \mathbf{y} &= \alpha_1(x^4 - 16) + \alpha_2(x^2 - x - 2) \\ &= \alpha_1 x^4 - \alpha_1(16) + \alpha_2 x^2 - \alpha_2 x - \alpha_2(2) \\ &= \alpha_1 x^4 + \alpha_2 x^2 - \alpha_2 x - \alpha_1(16) - \alpha_2(2) \end{aligned}$$

Now, equate coefficients to obtain the system of linear equations

$$\begin{aligned} 1 &= \alpha_1 \\ 2 &= \alpha_2 \\ -5 &= -\alpha_2 \\ -14 &= -16\alpha_1 - 2\alpha_2 \end{aligned}$$

We could build an augmented matrix and discover that this is an inconsistent system, but the second and third equations make it readily apparent that the system is inconsistent. The lack of a solution means the supposed linear combination is impossible and $\mathbf{y} \notin U$.⊚

Lets 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 [160]) and scalar matrix multiplication (Definition SMM [160]), as was show in Example VSM [241]. 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 [260]), $W = \mathcal{S}p(S)$. So by Theorem SSS [260] 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 $\mathbf{y}$ 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

$$\begin{aligned}
&\begin{bmatrix} 9 & 3 \\ 7 & 3 \\ 10 & -11 \end{bmatrix} \\
&= \alpha_1 \begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -5 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 14 & -1 \end{bmatrix} + \alpha_3 \begin{bmatrix} 3 & -1 \\ -1 & 2 \\ -19 & -11 \end{bmatrix} + \alpha_4 \begin{bmatrix} 4 & 2 \\ 1 & -2 \\ 14 & -2 \end{bmatrix} + \alpha_5 \begin{bmatrix} 3 & 1 \\ -4 & 0 \\ -17 & 7 \end{bmatrix} \\
&= \begin{bmatrix} 3\alpha_1 + \alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 & \alpha_1 + \alpha_2 - \alpha_3 + 2\alpha_4 + \alpha_5 \\ 4\alpha_1 + 2\alpha_2 - \alpha_3 + \alpha_4 - 4\alpha_5 & 2\alpha_1 - \alpha_2 + 2\alpha_3 - 2\alpha_4 \\ 5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 & -5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 \end{bmatrix}
\end{aligned}$$

Using our definition of matrix equality (Definition ME [159]) we can translate this statement into six equations in the five unknowns,

$$\begin{aligned}
3\alpha_1 + \alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 &= 9 \\
\alpha_1 + \alpha_2 - \alpha_3 + 2\alpha_4 + \alpha_5 &= 3 \\
4\alpha_1 + 2\alpha_2 - \alpha_3 + \alpha_4 - 4\alpha_5 &= 7 \\
2\alpha_1 - \alpha_2 + 2\alpha_3 - 2\alpha_4 &= 3 \\
5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 &= 10 \\
-5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 &= -11
\end{aligned}$$

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

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & \frac{5}{8} & 2 \\ 0 & \boxed{1} & 0 & 0 & \frac{-19}{4} & -1 \\ 0 & 0 & \boxed{1} & 0 & \frac{-7}{8} & 0 \\ 0 & 0 & 0 & \boxed{1} & \frac{17}{8} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So we recognize that the system is consistent since there is no leading 1 in the final column (Theorem RCLS [56]), and compute $n-r = 5-4 = 1$ free variables (Theorem FVCS [57]). While there are infinitely many solutions, we are only in pursuit of a single solution, so lets 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 $\mathbf{y}$, 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 $W = \mathcal{S}p(S)$. Second, is

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

in W? To answer this, we want to determine if $\mathbf{x}$ 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

$$\begin{aligned}
&\begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & -2 \end{bmatrix} \\
&= \alpha_1\begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 5 & -5 \end{bmatrix} + \alpha_2\begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 14 & -1 \end{bmatrix} + \alpha_3\begin{bmatrix} 3 & -1 \\ -1 & 2 \\ -19 & -11 \end{bmatrix} + \alpha_4\begin{bmatrix} 4 & 2 \\ 1 & -2 \\ 14 & -2 \end{bmatrix} + \alpha_5\begin{bmatrix} 3 & 1 \\ -4 & 0 \\ -17 & 7 \end{bmatrix} \\
&= \begin{bmatrix} 3\alpha_1 + \alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 & \alpha_1 + \alpha_2 - \alpha_3 + 2\alpha_4 + \alpha_5 \\ 4\alpha_1 + 2\alpha_2 - \alpha_3 + \alpha_4 - 4\alpha_5 & 2\alpha_1 - \alpha_2 + 2\alpha_3 - 2\alpha_4 \\ 5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 & -5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 \end{bmatrix}
\end{aligned}$$

Using our definition of matrix equality (Definition ME [159]) we can translate this statement into six equations in the five unknowns,

$$\begin{aligned}
3\alpha_1 + \alpha_2 + 3\alpha_3 + 4\alpha_4 + 3\alpha_5 &= 2 \\
\alpha_1 + \alpha_2 - \alpha_3 + 2\alpha_4 + \alpha_5 &= 1 \\
4\alpha_1 + 2\alpha_2 - \alpha_3 + \alpha_4 - 4\alpha_5 &= 3 \\
2\alpha_1 - \alpha_2 + 2\alpha_3 - 2\alpha_4 &= 1 \\
5\alpha_1 + 14\alpha_2 - 19\alpha_3 + 14\alpha_4 - 17\alpha_5 &= 4 \\
-5\alpha_1 - \alpha_2 - 11\alpha_3 - 2\alpha_4 + 7\alpha_5 &= -2
\end{aligned}$$

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

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & \frac{5}{8} & 0 \\ 0 & \boxed{1} & 0 & 0 & -\frac{38}{8} & 0 \\ 0 & 0 & \boxed{1} & 0 & -\frac{7}{8} & 0 \\ 0 & 0 & 0 & \boxed{1} & -\frac{17}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

With a leading 1 in the last column Theorem RCLS [56] 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 [261] and Example SM32 [262] 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 [159] are also subspaces — they are closed under vector addition and scalar multiplication in $\mathbb{C}^m$.

Theorem RMS
Range of a Matrix is a Subspace
Suppose that A is an $m \times n$ matrix. Then $\mathcal{R}(A)$ is a subspace of $\mathbb{C}^m$. □

Proof Definition RM [167] shows us that $\mathcal{R}(A)$ 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 $\mathcal{R}(A)$ is a span, Theorem SSS [260] 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 [121] 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 [258]. 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 $\nabla s(A)$ is a subspace of $\mathbb{C}^n$. □

Proof Theorem RMRST [188] applied to A^t, along with Theorem TT [165], yields

$$\nabla s(A) = \nabla s\left(\left(A^t\right)^t\right) = \mathcal{R}\left(A^t\right)$$

so the row space is equal to the range of a matrix, which Theorem RMS [265] says is a subspace. ■

So the span of a set of vectors, and the null space, range, and row space of a matrix are all subspaces, and hence are all vector spaces, meaning they have all the properties detailed in Section VS [239]. We have worked with these objects as just sets in Chapter V [89] and Chapter M [159], 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

$$\left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \,\middle|\, 3a - 2b + c = 5 \right\}$$

   Is this set a subspace of $\mathbb{C}^3$?
3. Name four general constructions of sets of vectors that we can now automatically deem as subspaces.

Subsection EXC
Exercises

C20 Contributed by Robert Beezer
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.

$$W = \mathcal{S}p\left(\left\{x^3 + x^2 + x,\, x^3 + 2x - 6,\, x^2 - 5\right\}\right)$$

Solution [269]

M20 Contributed by Robert Beezer
In $\mathbb{C}^3$, the vector space of column vectors of size 3, prove that the set Z is a subspace.

$$Z = \left\{\left.\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\,\right|\, 4x_1 - x_2 + 5x_3 = 0\right\}$$

Solution [269]

Subsection SOL
Solutions

C20 Exercise [267] Contributed by Robert Beezer
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 [242])

$$\begin{aligned} p(x) &= a_1(x^3 + x^2 + x) + a_2(x^3 + 2x - 6) + a_3(x^2 - 5) \\ x^3 + 6x + 4 &= (a_1 + a_2)x^3 + (a_1 + a_3)x^2 + (a_1 + 2a_2)x + (-6a_2 - 5a_3) \end{aligned}$$

Equating coefficients of equal powers of x, we get the system of equations,

$$\begin{aligned} a_1 + a_2 &= 1 \\ a_1 + a_3 &= 0 \\ a_1 + 2a_2 &= 6 \\ -6a_2 - 5a_3 &= 4 \end{aligned}$$

The augmented matrix of this system of equations row-reduces to

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

There is a leading 1 in the last column, so Theorem RCLS [56] implies that the system is inconsistent. So there is no way for p to gain membership in W, so $p \notin W$.

M20 Exercise [267] Contributed by Robert Beezer
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, $Z = \mathcal{N}\left(\begin{bmatrix} 4 & -1 & 5 \end{bmatrix}\right)$. Every null space is a subspace by Theorem NSMS [258].

A less direct solution appeals to Theorem TSS [255].

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 $\mathbf{0} \in Z$.

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 Z. Then we know that these vectors cannot be totally arbitrary, they must have gained membership in 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 $\mathbf{x} + \mathbf{y} \in Z$? 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,

$$\begin{aligned} & 4(x_1 + y_1) - 1(x_2 + y_2) + 4(x_3 + y_3) \\ & = 4x_1 + 4y_1 - x_2 - y_2 + 5x_3 + 5y_3 \\ & = (4x_1 - x_2 + 5x_3) + (4y_1 - y_2 + 5y_3) \\ & = 0 + 0 && \mathbf{x} \in Z,\ \mathbf{y} \in Z \\ & = 0 \end{aligned}$$

and by this computation we see that $\mathbf{x} + \mathbf{y} \in Z$.

If α is a scalar and $\mathbf{x} \in Z$, is it always true that $\alpha\mathbf{x} \in Z$? 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

$$\begin{aligned} 4(\alpha x_1) - (\alpha x_2) + 5(\alpha x_3) & \\ = \alpha(4x_1 - x_2 + 5x_3) & \\ = \alpha 0 & && \mathbf{x} \in Z \\ = 0 & \end{aligned}$$

and we see that indeed $\alpha\mathbf{x} \in Z$. With the three conditions of Theorem TSS [255] fulfilled, we can conclude that Z is a subspace of $\mathbb{C}^3$.

Section B
Bases

A basis of a vector space is one of the most useful concepts in linear algebra. It often provides a finite description of an infinite vector space. But before we can define a basis we need to return to the idea of linear independence.

Subsection LI
Linear independence

Our previous definition of linear independence (Definition LI [271]) employed a relation of linear dependence that had a linear combination on one side of the equality. As a linear combination in a vector space (Definition LC [259]) depends only on vector addition and scalar multiplication we can extend our definition of linear independence from the setting of $\mathbb{C}^m$ to the setting of a general vector space V. Compare these definitions with Definition RLDCV [129] and Definition LICV [129].

Definition RLD
Relation of Linear Dependence
Suppose that V is a vector space. Given a set of vectors $S = \{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_n\}$, an equation of the form

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \alpha_3\mathbf{u}_3 + \cdots + \alpha_n\mathbf{u}_n = \mathbf{0}$$

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. △

Definition LI
Linear Independence
Suppose that V is a vector space. The set of vectors $S = \{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_n\}$ 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. △

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 [242]) 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

$$\begin{aligned} &3\left(2x^4 + 3x^3 + 2x^2 - x + 10\right) + 4\left(-x^4 - 2x^3 + x^2 + 5x - 8\right) \\ &\quad + (-1)\left(2x^4 + x^3 + 10x^2 + 17x - 2\right) = 0x^4 + 0x^3 + 0x^2 + 0x + 0 = \mathbf{0}. \end{aligned}$$

This is a nontrivial relation of linear dependence (Definition RLD [271]) on the set S and so convinces us that S is linearly dependent (Definition LI [271]).

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. Lets look at another set of vectors (polynomials) from P_4. Let

$$\begin{aligned} T = \{&3x^4 - 2x^3 + 4x^2 + 6x - 1,\ -3x^4 + 1x^3 + 0x^2 + 4x + 2, \\ &4x^4 + 5x^3 - 2x^2 + 3x + 1,\ 2x^4 - 7x^3 + 4x^2 + 2x + 1\} \end{aligned}$$

Suppose we have a relation of linear dependence on this set,

$$\begin{aligned} \mathbf{0} = &\alpha_1\left(3x^4 - 2x^3 + 4x^2 + 6x - 1\right) + \alpha_2\left(-3x^4 + 1x^3 + 0x^2 + 4x + 2\right) \\ &+ \alpha_3\left(4x^4 + 5x^3 - 2x^2 + 3x + 1\right) + \alpha_4\left(2x^4 - 7x^3 + 4x^2 + 2x + 1\right) \end{aligned}$$

Using our definitions of vector addition and scalar multiplication in P_4 (Example VSP [242]), we arrive at,

$$\begin{aligned} 0x^4 + 0x^3 + 0x^2 + 0x + 0 = &\left(3\alpha_1 - 3\alpha_2 + 4\alpha_3 + 2\alpha_4\right)x^4 + \left(-2\alpha_1 + \alpha_2 + 5\alpha_3 - 7\alpha_4\right)x^3 \\ &+ \left(4\alpha_1 + -2\alpha_3 + 4\alpha_4\right)x^2 + \left(6\alpha_1 + 4\alpha_2 + 3\alpha_3 + 2\alpha_4\right)x \\ &+ \left(-\alpha_1 + 2\alpha_2 + \alpha_3 + \alpha_4\right). \end{aligned}$$

Equating coefficients, we arrive at the homogeneous system of equations,

$$\begin{aligned} 0 &= 3\alpha_1 - 3\alpha_2 + 4\alpha_3 + 2\alpha_4 \\ 0 &= -2\alpha_1 + \alpha_2 + 5\alpha_3 - 7\alpha_4 \\ 0 &= 4\alpha_1 + -2\alpha_3 + 4\alpha_4 \\ 0 &= 6\alpha_1 + 4\alpha_2 + 3\alpha_3 + 2\alpha_4 \\ 0 &= -\alpha_1 + 2\alpha_2 + \alpha_3 + \alpha_4 \end{aligned}$$

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

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We expected the system to be consistent (Theorem HSC [68]) and so can compute $n-r = 4-4 = 0$ and Theorem CSRN [57] tells us that the solution is unique. Since this is a homogeneous system, this unique solution is the trivial solution (Definition TSHSE [68]), $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$, $\alpha_4 = 0$. So by Definition LI [271] 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 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 [3] to provide this demonstration. Whew! There's a lot going on in this example. Spend some time with it, we'll still be 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 [241])

$$R = \left\{ \begin{bmatrix} 3 & -1 \\ 1 & 4 \\ 6 & -6 \end{bmatrix}, \begin{bmatrix} -2 & 3 \\ 1 & -3 \\ -2 & -6 \end{bmatrix}, \begin{bmatrix} 6 & -6 \\ -1 & 0 \\ 7 & -9 \end{bmatrix}, \begin{bmatrix} 7 & 9 \\ -4 & -5 \\ 2 & 5 \end{bmatrix} \right\}$$
$$S = \left\{ \begin{bmatrix} 2 & 0 \\ 1 & -1 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} -4 & 0 \\ -2 & 2 \\ -2 & -6 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -2 & 1 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} -5 & 3 \\ -10 & 7 \\ 2 & 0 \end{bmatrix} \right\}$$

One set is linearly independent, the other is not. Which is which? Lets examine R first. Build a generic relation of linear dependence,

$$\alpha_1 \begin{bmatrix} 3 & -1 \\ 1 & 4 \\ 6 & -6 \end{bmatrix} + \alpha_2 \begin{bmatrix} -2 & 3 \\ 1 & -3 \\ -2 & -6 \end{bmatrix} + \alpha_3 \begin{bmatrix} 6 & -6 \\ -1 & 0 \\ 7 & -9 \end{bmatrix} + \alpha_4 \begin{bmatrix} 7 & 9 \\ -4 & -5 \\ 2 & 5 \end{bmatrix} = \mathbf{0}$$

Massaging the left-hand side with our definitions of vector addition and scalar multiplication in M_{32} (Example VSM [241]) 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 [159]) and equating corresponding entries we get the homogeneous system of six equations in four variables,

$$\begin{aligned}
3\alpha_1 - 2\alpha_2 + 6\alpha_3 + 7\alpha_4 &= 0\\
-1\alpha_1 + 3\alpha_2 - 6\alpha_3 + 9\alpha_4 &= 0\\
1\alpha_1 + 1\alpha_2 - \alpha_3 - 4\alpha_4 &= 0\\
4\alpha_1 - 3\alpha_2 + -5\alpha_4 &= 0\\
6\alpha_1 - 2\alpha_2 + 7\alpha_3 + 2\alpha_4 &= 0\\
-6\alpha_1 - 6\alpha_2 - 9\alpha_3 + 5\alpha_4 &= 0
\end{aligned}$$

Form the augmented matrix of this homogeneous system and row-reduce to obtain

$$\begin{bmatrix}
\boxed{1} & 0 & 0 & 0 & 0\\
0 & \boxed{1} & 0 & 0 & 0\\
0 & 0 & \boxed{1} & 0 & 0\\
0 & 0 & 0 & \boxed{1} & 0\\
0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

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 [271]).

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 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 [241]) 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 [159]) and equating corresponding entries we get the homogeneous system of six equations in four variables,

$$\begin{aligned}
2\alpha_1 - 4\alpha_2 + \alpha_3 - 5\alpha_4 &= 0\\
+\alpha_3 + 3\alpha_4 &= 0\\
\alpha_1 - 2\alpha_2 - 2\alpha_3 - 10\alpha_4 &= 0\\
-\alpha_1 + 2\alpha_2 + \alpha_3 + 7\alpha_4 &= 0\\
\alpha_1 - 2\alpha_2 + 2\alpha_3 + 2\alpha_4 &= 0\\
3\alpha_1 - 6\alpha_2 + 4\alpha_3 &= 0
\end{aligned}$$

Form the augmented matrix of this homogeneous system and row-reduce to obtain

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

Analyzing this we see that the system is consistent (we expected that since the system is homogeneous, Theorem HSC [68]) 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 a linearly dependent set (Definition LI [271]). 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? ⊚

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, $\mathcal{S}p(S)$, using Definition SS [260] and then be assured by Theorem SSS [260] that the construction does provide a subspace. We now turn the situation upside-down. Suppose we are first given a subspace $W \subseteq V$. Can we find a set S so that $\mathcal{S}p(S) = W$? 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 TSS
To Span a Subspace
Suppose V is a vector space and W is a subspace. A subset S of W is a **spanning set** for W if $\mathcal{S}p(S) = W$. In this case, we also say S **spans** W. △

The definition of a spanning set requires that two sets (subspaces actually) be equal. If S is a subset of W, then $\mathcal{S}p(S) \subseteq W$, always. Thus it is usually only necessary to prove that $W \subseteq \mathcal{S}p(S)$. Now would be a good time to review Technique SE [21].

Example SSP4
Spanning set in P_4
In Example SP4 [256] we showed that

$$W = \{\, p(x) \mid p \in P_4,\ p(2) = 0\}$$

is a subspace of P_4, the vector space of polynomials with degree at most 4 (Example VSP [242]). 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 $W = \mathcal{S}p(S)$. 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, $\mathcal{S}p(S) \subseteq W$ also. So it remains to show that $W \subseteq \mathcal{S}p(S)$ (Technique SE [21]). So choose 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 $\mathcal{S}p(S)$, 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.

$$\begin{aligned}
r(x) &= ax^4 + bx^3 + cx^2 + dx + e \\
&= \alpha_1 (x - 2) + \alpha_2 \left(x^2 - 4x + 4\right) + \alpha_3 \left(x^3 - 6x^2 + 12x - 8\right) \\
&\quad + \alpha_4 \left(x^4 - 8x^3 + 24x^2 - 32x + 16\right) \\
&= \alpha_4 x^4 + (\alpha_3 - 8\alpha_4)\, x^3 + (\alpha_2 - 6\alpha_3 + 24\alpha_2)\, x^2 \\
&\quad + (\alpha_1 - 4\alpha_2 + 12\alpha_3 - 32\alpha_4)\, x + (-2\alpha_1 + 4\alpha_2 - 8\alpha_3 + 16\alpha_4)
\end{aligned}$$

Equating coefficients (vector equality in P_4) gives the system of five equations in four variables,

$$\begin{aligned}\alpha_4 &= a\\ \alpha_3 - 8\alpha_4 &= b\\ \alpha_2 - 6\alpha_3 + 24\alpha_2 &= c\\ \alpha_1 - 4\alpha_2 + 12\alpha_3 - 32\alpha_4 &= d\\ -2\alpha_1 + 4\alpha_2 - 8\alpha_3 + 16\alpha_4 &= e\end{aligned}$$

Any solution to this system of equations will provide the linear combination we need to determine if $r \in \mathcal{S}p(S)$, 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

$$\begin{bmatrix}\boxed{1} & 0 & 0 & 0 & 32a-12b+4c+d\\ 0 & \boxed{1} & 0 & 0 & -24a+6b+c\\ 0 & 0 & \boxed{1} & 0 & 8a+b\\ 0 & 0 & 0 & \boxed{1} & a\\ 0 & 0 & 0 & 0 & 16a+8b+4c+2d+e\end{bmatrix} = \begin{bmatrix}\boxed{1} & 0 & 0 & 0 & 32a-12b+4c+d\\ 0 & \boxed{1} & 0 & 0 & -24a+6b+c\\ 0 & 0 & \boxed{1} & 0 & 8a+b\\ 0 & 0 & 0 & \boxed{1} & a\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$$

where 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 [56] tells us this system is consistent. Therefore, $W \subseteq \mathcal{S}p(S)$ and so $W = \mathcal{S}p(S)$ and S is a spanning set for W.

Notice that an alternative to row-reducing the augmented matrix by hand would be to appeal to Theorem RNS [177] by expressing the range of the coefficient matrix as a null space, and then verifying that the condition on r guarantees that r is in the range, 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 [276] 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

$$Z = \left\{ \begin{bmatrix} a & b\\ c & d\end{bmatrix} \;\middle|\; a + 2b - 7d = 0,\ 3a - b + 7c - 7d = 0 \right\}$$

We will construct a limited number of matrices in Z and hope that they form a spanning set. We want to avoid any linear combinations of matrices we have already chosen as they will not provide us with anything new to build Z with. And the zero matrix should be avoided at all costs, it is of no help. Notice that if we choose a and b, the the first condition defining the set will determine the value d and then the second condition will determine c. This reminds me of how free variables behave, and so I am going to choose $a = 1$ and $b = 0$, and then $a = 0$ with $b = 1$. This creates two matrices

$$\begin{bmatrix} 1 & 0 \\ -\frac{2}{7} & \frac{1}{7} \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 \\ \frac{3}{7} & \frac{2}{7} \end{bmatrix}$$

Because it will be easier to work with integers, I will multiply each matrix by the scalar 7 (Definition SMM [160]), which should not change any of the properties relative to being a spanning set. The new matrices are then

$$Q = \left\{ \begin{bmatrix} 7 & 0 \\ -2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 7 \\ 3 & 2 \end{bmatrix} \right\}$$

The question is: if we take an arbitrary element of Z, can we write it as a linear combination of these two matrices in Q? Let's try.

$$\begin{aligned} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \alpha_1 \begin{bmatrix} 7 & 0 \\ -2 & 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 7 \\ 3 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 7\alpha_1 & 7\alpha_2 \\ -2\alpha_1 + 3\alpha_2 & \alpha_1 + 2\alpha_2 \end{bmatrix} \end{aligned}$$

Using our definition of matrix equality (Definition ME [159]) we equate corresponding entries and get a system of four equations in two variables,

$$\begin{aligned} 7\alpha_1 &= a \\ 7\alpha_2 &= b \\ -2\alpha_1 + 3\alpha_2 &= c \\ \alpha_1 + 2\alpha_2 &= d \end{aligned}$$

Form the augmented matrix and row-reduce (by hand),

$$\begin{bmatrix} \boxed{1} & 0 & \frac{a}{7} \\ 0 & \boxed{1} & \frac{b}{7} \\ 0 & 0 & \frac{1}{7}(2a - 3b + 7c) \\ 0 & 0 & \frac{1}{7}(a + 2b - 7d) \end{bmatrix}$$

In order to obtain a spanning set, we want this system to have a solution for any matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in Z$. However, this matrix is not as arbitrary as it appears, we may also assume (or require) that $a + 2b - 7d = 0$ and $3a - b + 7c - 7d = 0$. To be a consistent system, Theorem RCLS [56] tells us that we need to have no leading 1's in the final column of the row-reduced matrix. We see that the expression in the fourth row is

$$\frac{1}{7}(a + 2b - 7d) = \frac{1}{7}(0) = 0.$$

With a bit more finesse, the expression in the third row is

$$\begin{aligned}\frac{1}{7}(2a-3b+7c) &= \frac{1}{7}(2a-3b+7c+0)\\ &= \frac{1}{7}(2a-3b+7c+(a+2b-7d))\\ &= \frac{1}{7}(3a-b+7c-7d)\\ &= \frac{1}{7}(0)\\ &= 0.\end{aligned}$$

So the system is consistent, the desired scalars can be found as a solution in every event, so $Z \subseteq \mathcal{S}p(Q)$ and Q is a spanning set for Z. ⊚

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. △

So, a basis is a linearly independent spanning set for a vector space. The requirement that the set spans 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 [293], 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 [138], Theorem BROC [172], Theorem BRS [186]) 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 [139], Example ROCD [173] and Example IAS [187]. 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 range, 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 [279] 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 range 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 $\mathbb{C}^m$, $B = \{\mathbf{e}_1,\, \mathbf{e}_2,\, \mathbf{e}_3,\, \ldots,\, \mathbf{e}_m\} = \{\mathbf{e}_i \mid 1 \leq i \leq m\}$ 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 [218], the columns of the identity matrix, which we know is nonsingular (since it row-reduces to the identity matrix!). And the columns of a nonsingular matrix are linearly independent by Theorem NSLIC [137].

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.

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix} = v_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + v_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + v_m \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3 + \cdots + v_m\mathbf{e}_m$$

this shows that $\mathbb{C}^m \subseteq \mathcal{S}p(B)$, 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

$$B = \left\{1,\, x,\, x^2,\, x^3,\, \ldots,\, x^n\right\}.$$

Another nice basis for P_n is

$$C = \left\{1,\, 1 + x,\, 1 + x + x^2,\, 1 + x + x^2 + x^3,\, \ldots,\, 1 + x + x^2 + x^3 + \cdots + x^n\right\}.$$

Checking that each of B and C is a linearly independent spanning set are good exercises. ⊙

Example BM
A basis for the vector space of matrices
In the vector space M_{mn} of matrices (Example VSM [241]) define the matrices $B_{k\ell}$, $1 \leq k \leq m$, $1 \leq \ell \leq n$ 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,

$$B = \{\, B_{k\ell} \mid 1 \leq k \leq m,\ 1 \leq \ell \leq n \}$$

forms a basis for M_{mn}. ⊙

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 [276] 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 $W = \{p(x) \mid p \in P_4,\ p(2) = 0\}$. We will now show that S is also linearly independent in W. Begin with a relation of linear dependence,

$$\begin{aligned} 0 + 0x + 0x^2 + 0x^3 + 0x^4 &= \alpha_1 (x - 2) + \alpha_2 \left(x^2 - 4x + 4\right) \\ &\quad + \alpha_3 \left(x^3 - 6x^2 + 12x - 8\right) + \alpha_4 \left(x^4 - 8x^3 + 24x^2 - 32x + 16\right) \\ &= \alpha_4 x^4 + (\alpha_3 - 8\alpha_4)\, x^3 + (\alpha_2 - 6\alpha_3 + 24\alpha_4)\, x^2 \\ &\quad + (\alpha_1 - 4\alpha_2 + 12\alpha_3 - 32\alpha_4)\, x + (-2\alpha_1 + 4\alpha_2 - 8\alpha_3 + 16\alpha_4) \end{aligned}$$

Equating coefficients (vector equality in P_4) gives the homogeneous system of five equations in four variables,

$$\begin{aligned} \alpha_4 &= 0 \\ \alpha_3 - 8\alpha_4 &= 0 \\ \alpha_2 - 6\alpha_3 + 24\alpha_4 &= 0 \\ \alpha_1 - 4\alpha_2 + 12\alpha_3 - 32\alpha_4 &= 0 \\ -2\alpha_1 + 4\alpha_2 - 8\alpha_3 + 16\alpha_4 &= 0 \end{aligned}$$

We form the augmented matrix, and row-reduce to obtain a matrix in reduced row-echelon form

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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. Finally, S has earned the right to be called a basis for W. ⊚

Example BSM22
A basis for a subspace of M_{22}
In Example SSM22 [277] we discovered that

$$Q = \left\{ \begin{bmatrix} 7 & 0 \\ -2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 7 \\ 3 & 2 \end{bmatrix} \right\}$$

is a spanning set for the subspace

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle|\ a + 2b - 7d = 0,\ 3a - b + 7c - 7d = 0 \right\}$$

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.

$$\begin{aligned} \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} &= \alpha_1 \begin{bmatrix} 7 & 0 \\ -2 & 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 7 \\ 3 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 7\alpha_1 & 7\alpha_2 \\ -2\alpha_1 + 3\alpha_2 & \alpha_1 + 2\alpha_2 \end{bmatrix} \end{aligned}$$

Using our definition of matrix equality (Definition ME [159]) we equate corresponding entries and get a homogeneous system of four equations in two variables,

$$\begin{aligned} 7\alpha_1 &= 0 \\ 7\alpha_2 &= 0 \\ -2\alpha_1 + 3\alpha_2 &= 0 \\ \alpha_1 + 2\alpha_2 &= 0 \end{aligned}$$

We could row-reduce the augmented matrix of this system, but it is not necessary. The first two 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. Therefore Q is a basis for Z. ⊚

Subsection BRS
Bases from Row Spaces

We have seen several examples of bases in different vector spaces. In this subsection, and the next (Subsection B.BNSM [285]), 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 [185] says that row-equivalent matrices have identical row spaces, while Theorem BRS [186] 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 vectors, in Example SCAD [121] we looked at the set

$$W = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\-3\\1\end{bmatrix},\ \begin{bmatrix}1\\4\\1\end{bmatrix},\ \begin{bmatrix}7\\-5\\4\end{bmatrix},\ \begin{bmatrix}-7\\-6\\-5\end{bmatrix}\right\}\right)$$

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 [183], the row space of C will be W, $\nabla s(C) = W$. Theorem BRS [186] 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 $\nabla s(C)$, and hence a basis for W. Let's do it — C row-reduces to

$$\begin{bmatrix}\boxed{1} & 0 & \frac{7}{11}\\ 0 & \boxed{1} & \frac{1}{11}\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

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

$$W = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\\frac{7}{11}\end{bmatrix},\ \begin{bmatrix}0\\1\\\frac{1}{11}\end{bmatrix}\right\}\right)$$

⊚

Example IAS [187] 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 [134] we began with a set of $n = 4$ vectors from $\mathbb{C}^5$,

$$R = \{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\mathbf{v}_4\} = \left\{ \begin{bmatrix}1\\2\\-1\\3\\2\end{bmatrix},\ \begin{bmatrix}2\\1\\3\\1\\2\end{bmatrix},\ \begin{bmatrix}0\\-7\\6\\-11\\-2\end{bmatrix},\ \begin{bmatrix}4\\1\\2\\1\\6\end{bmatrix} \right\}$$

and defined $V = \mathcal{S}p(R)$. 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 $\mathcal{S}p(R)$. By Theorem BRS [186] 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

$$\left\{ \begin{bmatrix}1\\0\\0\\-\frac{1}{17}\\\frac{30}{17}\end{bmatrix},\ \begin{bmatrix}0\\1\\0\\\frac{25}{17}\\-\frac{2}{17}\end{bmatrix},\ \begin{bmatrix}0\\0\\1\\-\frac{2}{17}\\-\frac{8}{17}\end{bmatrix} \right\}$$

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 BNSM
Bases and NonSingular Matrices

A quick source of diverse bases for $\mathbb{C}^m$ is the set of columns of a nonsingular matrix.

Theorem CNSMB
Columns of NonSingular Matrix are a Basis
Suppose that A is a square matrix. Then the columns of A are a basis of $\mathbb{C}^m$ if and only if A is nonsingular. □

Proof (⇒) Suppose that the columns of A are a basis for $\mathbb{C}^m$. Then Definition B [279] says the set of columns is linearly independent. Theorem NSLIC [137] then says that A is nonsingular.

(⇐) Suppose that A is nonsingular. Then by Theorem NSLIC [137] this set of columns is linearly independent. Theorem RNSM [180] says that for a nonsingular matrix, $\mathcal{R}(A) = \mathbb{C}^m$. 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 [279]). ■

Example CABAK
Columns as Basis, Archetype K
Archetype K [520] 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 NSRRI [76], K is nonsingular. Then Theorem CNSMB [285] 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 [280]) as just a simple corollary of Theorem CNSMB [285]?

With a new equivalence for a nonsingular matrix, we can update our list of equivalences (Theorem NSME4 [229]).

Theorem NSME5
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. The range of A is $\mathbb{C}^n$, $\mathcal{R}(A) = \mathbb{C}^n$.
7. A is invertible.
8. The columns of A are a basis for $\mathbb{C}^n$. □

Subsection VR
Vector Representation

In Chapter R [441] we will take up the matter of representations fully. Now we will 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 a basis as a minimal spanning set. First an example, then the theorem.

Example AVR
A vector representation
The set

$$\left\{ \begin{bmatrix} -7 \\ 5 \\ 1 \end{bmatrix},\, \begin{bmatrix} -6 \\ 5 \\ 0 \end{bmatrix},\, \begin{bmatrix} -12 \\ 7 \\ 4 \end{bmatrix} \right\}$$

is a basis for $\mathbb{C}^3$. (You have the tools to check this right now, but this fact can be assumed for the purpose of this example.) This set comes from the columns of the coefficient matrix of Archetype B [478]. Because $\mathbf{x} = \begin{bmatrix} -3 \\ 5 \\ 2 \end{bmatrix}$ is a solution to this system, we can use Theorem SLSLC [101]

$$\begin{bmatrix} -33 \\ 24 \\ 5 \end{bmatrix} = (-3) \begin{bmatrix} -7 \\ 5 \\ 1 \end{bmatrix} + (5) \begin{bmatrix} -6 \\ 5 \\ 0 \end{bmatrix} + (2) \begin{bmatrix} -12 \\ 7 \\ 4 \end{bmatrix}.$$

Further, we know this is the *only* way we can express $\begin{bmatrix} -33 \\ 24 \\ 5 \end{bmatrix}$ as a linear combination of these basis vectors, since the nonsingularity of the coefficient matrix tells that this solution is unique. This is all an illustration of the following theorem. ⊚

Theorem VRRB
Vector Representation Relative to a Basis
Suppose that V is a vector space with basis $B = \{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3,\, \ldots,\, \mathbf{v}_m\}$ and that $\mathbf{w}$ is a vector in V. Then there exist *unique* scalars $a_1,\, a_2,\, a_3,\, \ldots,\, a_m$ such that

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_m\mathbf{v}_m.$$ □

Proof That $\mathbf{w}$ can be written as a linear combination of the basis vectors follows from the spanning property of the basis (Definition B [279], Definition TSS [276]). 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 [78])

Assume there are two ways to express $\mathbf{w}$ 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

$$\begin{aligned}\mathbf{w} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_m\mathbf{v}_m\\ \mathbf{w} &= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + \cdots + b_m\mathbf{v}_m.\end{aligned}$$

Then notice that (using the vector space axioms of associativity and distributivity)

$$\begin{aligned}\mathbf{0} &= \mathbf{w} - \mathbf{w}\\ &= (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_m\mathbf{v}_m) - (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + \cdots + b_m\mathbf{v}_m)\\ &= (a_1 - b_1)\mathbf{v_1} + (a_2 - b_2)\mathbf{v_2} + (a_3 - b_3)\mathbf{v_3} + \cdots + (a_m - b_m)\mathbf{v_m}\end{aligned}$$

But this is a relation of linear dependence on a linearly independent set of vectors! Now we are the other half of the assumption that $\{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3,\, \ldots,\, \mathbf{v}_m\}$ is a basis (Definition B [279]). So by Definition LI [271] it *must* happen that the scalars are all zero. That is,

$$\begin{array}{ccccc}(a_1 - b_1) = 0 & (a_2 - b_2) = 0 & (a_3 - b_3) = 0 & \ldots & (a_m - b_m) = 0\\ a_1 = b_1 & a_2 = b_2 & a_3 = b_3 & \ldots & a_m = b_m.\end{array}$$

And so we find that the scalars are unique. ■

This is a very typical use of the hypothesis that a set is linear 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 basis vectors, but only just. There is only enough raw material in the spanning set to write each vector one way as a linear combination. This theorem will be the basis (pun intended) for our future definition of coordinate vectors in Definition XXX [**??**].

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. The matrix below is nonsingular. What can you now say about its columns?

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

3. 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?

Subsection EXC
Exercises

C20 Contributed by Robert Beezer

In the vector space of 2×2 matrices, M_{22}, determine if the set S below is linearly independent.

$$S = \left\{ \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 4 \\ -1 & 2 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \right\}$$

Solution [291]

Subsection SOL
Solutions

C20 Exercise [289] Contributed by Robert Beezer
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 [159]) we arrive at a homogeneous system of linear equations,

$$\begin{aligned} 2a_1 + 4a_3 &= 0 \\ -a_1 + 4a_2 + 2a_3 &= 0 \\ a_1 - a_2 + a_3 &= 0 \\ 3a_1 + 2a_2 + 3a_3 &= 0 \end{aligned}$$

The coefficient matrix of this system row-reduces to the matrix,

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \end{bmatrix}$$

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 [271]) is trivial, the set S is linearly independent (Definition LI [271]).

Section D
Dimension

Almost every vector space we have encountered has been infinite in size (an exception is Example VSS [243]). 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 $\{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\ldots,\,\mathbf{v}_t\}$ 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 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 [293] 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. Here is a fundamental result that many subsequent theorems will trace their lineage back to.

Theorem SSLD
Spanning Sets and Linear Dependence
Suppose that $S = \{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\ldots,\,\mathbf{v}_t\}$ 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. □

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 $\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_m$ can be written as a linear combination of $\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\ldots,\,\mathbf{v}_t$ since S is a spanning set of V. This means there exist scalars a_{ij}, $1 \leq i \leq t$, $1 \leq j \leq m$, so that

$$\begin{aligned}
\mathbf{u}_1 &= a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3 + \cdots + a_{t1}\mathbf{v}_t\\
\mathbf{u}_2 &= a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + a_{32}\mathbf{v}_3 + \cdots + a_{t2}\mathbf{v}_t\\
\mathbf{u}_3 &= a_{13}\mathbf{v}_1 + a_{23}\mathbf{v}_2 + a_{33}\mathbf{v}_3 + \cdots + a_{t3}\mathbf{v}_t\\
&\vdots\\
\mathbf{u}_m &= a_{1m}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + a_{3m}\mathbf{v}_3 + \cdots + a_{tm}\mathbf{v}_t
\end{aligned}$$

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},

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1m}x_m &= 0\\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2m}x_m &= 0\\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3m}x_m &= 0\\
&\vdots\\
a_{t1}x_1 + a_{t2}x_2 + a_{t3}x_3 + \cdots + a_{tm}x_m &= 0
\end{aligned}$$

This is a homogeneous system with more variables than equations (our hypothesis is expressed as $m > t$), so by Theorem HMVEI [69] 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

$$\begin{aligned}
a_{11}c_1 + a_{12}c_2 + a_{13}c_3 + \cdots + a_{1m}c_m &= 0\\
a_{21}c_1 + a_{22}c_2 + a_{23}c_3 + \cdots + a_{2m}c_m &= 0\\
a_{31}c_1 + a_{32}c_2 + a_{33}c_3 + \cdots + a_{3m}c_m &= 0\\
&\vdots\\
a_{t1}c_1 + a_{t2}c_2 + a_{t3}c_3 + \cdots + a_{tm}c_m &= 0
\end{aligned}$$

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,

$$\begin{aligned}
&c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 + \cdots + c_m\mathbf{u}_m \\
&= c_1\left(a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2 + a_{31}\mathbf{v}_3 + \cdots + a_{t1}\mathbf{v}_t\right) \\
&\qquad + c_2\left(a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2 + a_{32}\mathbf{v}_3 + \cdots + a_{t2}\mathbf{v}_t\right) \\
&\qquad + c_3\left(a_{13}\mathbf{v}_1 + a_{23}\mathbf{v}_2 + a_{33}\mathbf{v}_3 + \cdots + a_{t3}\mathbf{v}_t\right) \\
&\qquad\qquad \vdots \\
&\qquad + c_m\left(a_{1m}\mathbf{v}_1 + a_{2m}\mathbf{v}_2 + a_{3m}\mathbf{v}_3 + \cdots + a_{tm}\mathbf{v}_t\right) \\
&= c_1a_{11}\mathbf{v}_1 + c_1a_{21}\mathbf{v}_2 + c_1a_{31}\mathbf{v}_3 + \cdots + c_1a_{t1}\mathbf{v}_t \\
&\qquad + c_2a_{12}\mathbf{v}_1 + c_2a_{22}\mathbf{v}_2 + c_2a_{32}\mathbf{v}_3 + \cdots + c_2a_{t2}\mathbf{v}_t \\
&\qquad + c_3a_{13}\mathbf{v}_1 + c_3a_{23}\mathbf{v}_2 + c_3a_{33}\mathbf{v}_3 + \cdots + c_3a_{t3}\mathbf{v}_t \\
&\qquad\qquad \vdots \\
&\qquad + c_ma_{1m}\mathbf{v}_1 + c_ma_{2m}\mathbf{v}_2 + c_ma_{3m}\mathbf{v}_3 + \cdots + c_ma_{tm}\mathbf{v}_t \\
&= \left(c_1a_{11} + c_2a_{12} + c_3a_{13} + \cdots + c_ma_{1m}\right)\mathbf{v}_1 \\
&\qquad + \left(c_1a_{21} + c_2a_{22} + c_3a_{23} + \cdots + c_ma_{2m}\right)\mathbf{v}_2 \\
&\qquad + \left(c_1a_{31} + c_2a_{32} + c_3a_{33} + \cdots + c_ma_{3m}\right)\mathbf{v}_3 \\
&\qquad\qquad \vdots \\
&\qquad + \left(c_1a_{t1} + c_2a_{t2} + c_3a_{t3} + \cdots + c_ma_{tm}\right)\mathbf{v}_t \\
&= \left(a_{11}c_1 + a_{12}c_2 + a_{13}c_3 + \cdots + a_{1m}c_m\right)\mathbf{v}_1 \\
&\qquad + \left(a_{21}c_1 + a_{22}c_2 + a_{23}c_3 + \cdots + a_{2m}c_m\right)\mathbf{v}_2 \\
&\qquad + \left(a_{31}c_1 + a_{32}c_2 + a_{33}c_3 + \cdots + a_{3m}c_m\right)\mathbf{v}_3 \\
&\qquad\qquad \vdots \\
&\qquad + \left(a_{t1}c_1 + a_{t2}c_2 + a_{t3}c_3 + \cdots + a_{tm}c_m\right)\mathbf{v}_t \\
&= 0\mathbf{v}_1 + 0\mathbf{v}_2 + 0\mathbf{v}_3 + \cdots + 0\mathbf{v}_t \\
&= \mathbf{0} + \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} \\
&= \mathbf{0}
\end{aligned}$$

That does it. R has been undeniably shown to be a linearly dependent set. ■

The proof just given has some rather 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 hopefully 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, $R = \{\mathbf{u}_j \mid 1 \leq j \leq m\}$, 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

$$\mathbf{u}_j = \sum_{i=1}^{t} a_{ij}\mathbf{v_i} \qquad\qquad 1 \leq j \leq m$$

Now we form, unmotivated, the homogeneous system of t equations in the m variables, x_j, $1 \leq j \leq m$, where the coefficients are the just-discovered scalars a_{ij},

$$\sum_{j=1}^{m} a_{ij}x_j = 0 \qquad\qquad 1 \leq i \leq t$$

This is a homogeneous system with more variables than equations (our hypothesis is expressed as $m > t$), so by Theorem HMVEI [69] there are infinitely many solutions. Choose one of these solutions that is not trivial and denote it by $x_j = c_j$, $1 \leq j \leq m$. As a solution to the homogeneous system, we then have $\sum_{j=1}^{m} a_{ij}c_j = 0$ for $1 \leq i \leq t$. As a collection of nontrivial scalars, c_j, $1 \leq j \leq m$, will provide the nontrivial relation of linear dependence we desire,

$$\begin{aligned}
\sum_{j=1}^{m} c_j\mathbf{u}_j &= \sum_{j=1}^{m} c_j \left(\sum_{i=1}^{t} a_{ij}\mathbf{v_i}\right) && \text{Substitution for } \mathbf{u}_j \\
&= \sum_{j=1}^{m}\sum_{i=1}^{t} c_j a_{ij}\mathbf{v}_i && \text{Distributivity (scalar)} \\
&= \sum_{i=1}^{t}\sum_{j=1}^{m} c_j a_{ij}\mathbf{v}_i && \text{Commutativity (vector addition)} \\
&= \sum_{i=1}^{t}\sum_{j=1}^{m} a_{ij} c_j\mathbf{v}_i && \text{Commutativity (scalar)} \\
&= \sum_{i=1}^{t}\left(\sum_{j=1}^{m} a_{ij} c_j\right)\mathbf{v}_i && \text{Distributivity (scalar multiplication)} \\
&= \sum_{i=1}^{t} 0\mathbf{v}_i && \text{Solution to homogeneous system} \\
&= \sum_{i=1}^{t} \mathbf{0} && \text{Theorem ZSSM [247]} \\
&= \mathbf{0} && \text{Property of the zero vector}
\end{aligned}$$

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 (. . .).

Theorem SSLD [293] can be viewed as a generalization of Theorem MVSLD [133]. We know that $\mathbb{C}^m$ has a basis with m vectors in it (Theorem SUVB [280]), so it is a set of m vectors that spans $\mathbb{C}^m$. By Theorem SSLD [293], 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 [133]. Maybe this is not a total shock, as the proofs of both theorems rely heavily on Theorem HMVEI [69]. The beauty of Theorem SSLD [293] 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 [276] 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 $W = \{\,p(x) \mid p \in P_4,\ p(2)=0\}$. So we can apply Theorem SSLD [293] 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,

$$T = \{p_1,\, p_2,\, p_3,\, p_4,\, p_5\} \subseteq W$$

$$\begin{aligned}
p_1 &= x^4 - 2x^3 + 2x^2 - 8x + 8\\
p_2 &= -x^3 + 6x^2 - 5x - 6\\
p_3 &= 2x^4 - 5x^3 + 5x^2 - 7x + 2\\
p_4 &= -x^4 + 4x^3 - 7x^2 + 6x\\
p_5 &= 4x^3 - 9x^2 + 5x - 6
\end{aligned}$$

By Theorem SSLD [293] we conclude that T is linearly dependent with no further computations. ⊚

Theorem SSLD [293] is indeed powerful, but our main purpose in proving it right now was to make sure that our definition of dimension (Definition D [293]) 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. □

Proof Suppose that C has more vectors than B. (Allowing for the possibility that C is infinite, we can replace C by a subset of that has more vectors than B.) As a basis, B is a spanning set for V (Definition B [279]), so Theorem SSLD [293] says that C is linearly dependent. However, this contradicts the fact that as a basis C is linearly independent (Definition B [279]). 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 [279]), so Theorem SSLD [293] says that B is linearly dependent. However, this contradicts the fact that as a basis B is linearly independent (Definition B [279]). 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 [297] 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 [293] 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 VSCM [241]) is m. □

Proof Theorem SUVB [280] provides a basis with m vectors. ■

Theorem DP
Dimension of P_n
The dimension of P_n (Example VSP [242]) is $n+1$. □

Proof Example BP [280] provides *two* bases with $n+1$ vectors. Take your pick. ■

Theorem DM
Dimension of M_{mn}
The dimension of M_{mn} (Example VSM [241]) is mn. □

Proof Example BM [281] provides a basis with mn vectors. ■

Example DSM22
Dimension of a subspace of M_{22}
It should now be plausible that

$$Z = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle|\, 2a + b + 3c + 4d = 0,\ -a + 3b - 5c - d = 0 \right\}$$

is a subspace of the vector space M_{22} (Example VSM [241]). (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

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

We can row-reduce this matrix to obtain

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

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,

$$\begin{aligned} a &= -2c - 2d \\ b &= c \end{aligned}$$

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

$$Z = \mathcal{S}p(S) = \mathcal{S}p\left(\left\{ \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix} \right\} \right)$$

Are these two matrices (vectors) also linearly independent? Begin with a relation of linear dependence on S,

$$\begin{aligned} a_1 \begin{bmatrix} -2 & 1 \\ 1 & 0 \end{bmatrix} + a_2 \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix} &= \mathcal{O} \\ \begin{bmatrix} -2a_1 - 2a_2 & a_1 \\ a_1 & a_2 \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

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 [271]). So S is a basis for V (Definition B [279]). Finally, we can conclude that $\dim(Z) = 2$ (Definition D [293]) since S has two elements.⊚

Example DSP4
Dimension of a subspace of P_4
In Example BSP4 [281] 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 $W = \{p(x) \mid p \in P_4,\ p(2) = 0\}$. Thus, the dimension of W is four, $\dim(W) = 4$. ⊚

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 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

$$P = \{p \mid p(x) \text{ is a polynomial in } x\}$$

Our operations will be the same as those defined for P_n (Example VSP [242]).

With no restrictions on the possible degrees of our polynomials, any finite set that is a candidate for spanning P will come up short. Suppose S is a potential finite spanning set and let m be the maximum degree of all the polynomials in S. If q is a polynomial of degree $m + 1$, then it will be impossible to form a linear combination that equals q by using elements of S. So no finite set can span P and no finite bases exist. Thus, $\dim(P) = \infty$. ⊚

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 [258]), the range (Theorem RMS [265]) and the row space (Theorem RSMS [265]). As vector spaces, each of these has a dimension, and for the null space and range, 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, $n(A) = \dim(\mathcal{N}(A))$. △

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 range of A, $r(A) = \dim(\mathcal{R}(A))$. △

Example RNM
Rank and nullity of a matrix

Let's compute the rank and nullity of

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

To do this, we will first row-reduce the matrix since that will help us determine bases for the null space and range.

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

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 BROC [172] creates a single basis vector. In total the basis will have 4 vectors, so the range of A will have dimension 4 and we write $r(A) = 4$.

For each index in F, Theorem BNS [138] 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 BROC [172] provides a basis for the range 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. So there are r column vectors in a basis for $\mathcal{R}(A)$.

Theorem BNS [138] 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 (Chapter A [469]) that involves a matrix lists its rank and nullity. You may have noticed as you studied the archetypes that the larger the range is the smaller the null space is. A simple corollary states this trade-off succinctly.

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 row-echelon form. By Theorem CRN [301],

$$r(A) + n(A) = r + (n - r) = n$$ ■

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"!

Subsection RNNSM
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,

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

With $n = 7$ columns and $r = 7$ nonzero rows Theorem CRN [301] 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 RNNSM
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, $r\left(A\right) = n$.
3. The nullity of A is zero, $n\left(A\right) = 0$. □

Proof ($1 \Rightarrow 2$) Theorem RNSM [180] says that if A is nonsingular then $\mathcal{R}(A) = \mathbb{C}^n$. If $\mathcal{R}(A) = \mathbb{C}^n$, then the range has dimension n by Theorem DCM [298], so the rank of A is n.
($2 \Rightarrow 3$) Suppose $r\left(A\right) = n$. Then Theorem RPNC [302] gives

$$n = r\left(A\right) + n\left(A\right) = n + n\left(A\right) \quad \Rightarrow \quad n\left(A\right) = 0$$

($3 \Rightarrow 1$) Suppose $n\left(A\right) = 0$, so a basis for the null space of A is the empty set. This implies that $\mathcal{N}(A) = \{\mathbf{0}\}$ and Theorem NSTNS [78] says A is nonsingular. ■

With a new equivalence for a nonsingular matrix, we can update our list of equivalences (Theorem NSME4 [229]) which now becomes a list requiring into double digits to number.

Theorem NSME6
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. The range of A is $\mathbb{C}^n$, $\mathcal{R}(A) = \mathbb{C}^n$.
7. A is invertible.
8. The columns of A are a basis for $\mathbb{C}^n$.
9. The rank of A is n, $r\left(A\right) = n$.
10. The nullity of A is zero, $n\left(A\right) = 0$. □

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?
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

M20 Contributed by Robert Beezer
M_{22} is the vector space of 2×2 matrices. Let S_{22} denote the set of all 2×2 symmetric matrices. That is

$$S_{22} = \left\{ A \in M_{22} \,\middle|\, A^t = A \right\}$$

(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}? Solution [307]

Subsection SOL
Solutions

M20 Exercise [305] Contributed by Robert Beezer
(a) We will use the three criteria of Theorem TSS [255]. The zero vector of M_{22} is the zero matrix, $\mathcal{O}$ (Definition ZM [163]), 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,

$$\begin{aligned} (A+B)^t &= A^t + B^t && \text{Theorem TASM [164]} \\ &= A + B && A,\, B \in S_{22} \end{aligned}$$

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,

$$\begin{aligned} \alpha A^t &= \alpha A^t && \text{Theorem TT [165]} \\ &= \alpha A && A \in S_{22} \end{aligned}$$

So αA is also symmetric and qualifies for membership in S_{22}.

With the three criteria of Theorem TSS [255] 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

$$\begin{aligned} \begin{bmatrix} a & b \\ b & d \end{bmatrix} &= \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ b & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix} \\ &= a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

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,

$$\begin{aligned} \mathcal{O} &= a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} &= \begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix} \end{aligned}$$

The equality of these two matrices (Definition ME [159]) 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 [293], $\dim(S_{22}) = 3$.

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 range 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 simply by adding vectors from outside the span of the linearly independent set and the linear independence of the set is preserved.

Theorem ELIS
Extending Linearly Independent Sets
Suppose V is vector space and S is a linearly independent set of vectors from V. Suppose $\mathbf{w}$ is a vector such that $\mathbf{w} \not\in \mathcal{S}p(S)$. Then the set $S' = S \cup \{\mathbf{w}\}$ is linearly independent.□

Proof Suppose $S = \{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\ldots,\,\mathbf{v}_m\}$ and begin with a relation of linear dependence on S',

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_m\mathbf{v}_m + a_{m+1}\mathbf{w} = \mathbf{0}.$$

There are two cases to consider. First suppose that $a_{m+1} = 0$. Then the relation of linear dependence on S' becomes

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_m\mathbf{v}_m = \mathbf{0}.$$

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

$$\begin{aligned} a_{m+1}\mathbf{w} &= -a_1\mathbf{v}_1 - a_2\mathbf{v}_2 - a_3\mathbf{v}_3 - \cdots - a_m\mathbf{v}_m \\ \mathbf{w} &= -\frac{a_1}{a_{m+1}}\mathbf{v}_1 - \frac{a_2}{a_{m+1}}\mathbf{v}_2 - \frac{a_3}{a_{m+1}}\mathbf{v}_3 - \cdots - \frac{a_m}{a_{m+1}}\mathbf{v}_m \end{aligned}$$

This equation expresses $\mathbf{w}$ as a linear combination of the vectors in S, contrary to the assumption that $\mathbf{w} \not\in \mathcal{S}p(S)$, 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, 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 $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_m\}$ 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 [279] and Theorem BIS [297] 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 [293].
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 [293].
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 $\mathbf{w} \notin \mathcal{S}p(S)$. By Theorem ELIS [309], 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 [293].
4. Suppose to the contrary that S is linearly dependent. Then by Theorem DLDS [134] (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 $V = \mathcal{S}p(S) = \mathcal{S}p(S')$. 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 [293]. ■

There is a tension in the construction of basis. Make the set too big and you will end up with relations of linear dependence among the vectors. Make the set too small and you will not have enough raw material to span the entire vector space. Make it 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 [310].

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 DLDS [134], so in a way we could think of this entire theorem as a corollary of Theorem DLDS [134]. 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 [309] and Theorem DLDS [134] before contradicting Theorem DLDS [134].

Theorem G [310] 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 [280] we claimed that

$$B = \left\{1,\, x,\, x^2,\, x^3,\, \ldots,\, x^n\right\}$$
$$C = \left\{1,\, 1+x,\, 1+x+x^2,\, 1+x+x^2+x^3,\, \ldots,\, 1+x+x^2+x^3+\cdots+x^n\right\}.$$

were both bases for P_n (Example VSP [242]). 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 [310] 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 [298] 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 [241]) given by

$$Z = \left\{\left.\begin{bmatrix}a & b\\ c & d\end{bmatrix}\,\right|\, 2a+b+3c+4d=0,\, -a+3b-5c-d=0\right\}$$

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.

$$2\begin{bmatrix}-2 & 1\\ 1 & 0\end{bmatrix} + (-3)\begin{bmatrix}-2 & 0\\ 0 & 1\end{bmatrix} = \begin{bmatrix}2 & 2\\ 2 & -3\end{bmatrix}$$
$$3\begin{bmatrix}-2 & 1\\ 1 & 0\end{bmatrix} + \begin{bmatrix}-2 & 0\\ 0 & 1\end{bmatrix} = \begin{bmatrix}-8 & 3\\ 3 & 1\end{bmatrix}$$

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

$$\begin{bmatrix} \boxed{1} & 0 \\ 0 & \boxed{1} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

So with $a_1 = a_2 = 0$ as the only solution, the set is linearly independent. Now we can apply Theorem G [310] to see that C also spans Z and therefore is a second basis for Z. ⊚

Example SVP4
Sets of vectors in P_4
In Example BSP4 [281] 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 $W = \{p(x) \mid p \in P_4,\ p(2) = 0\}$. So $\dim(W) = 4$.

The set

$$\left\{ 3x^2 - 5x - 2,\ 2x^2 - 7x + 6,\ x^3 - 2x^2 + x - 2 \right\}$$

is a subset of W (check this) and it happens to be linearly independent (check this, too). However, by Theorem G [310] 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 [310] 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 [310] tells us that it also spans W, and therefore is a basis of W.

Subsection RT
Ranks and Transposes

With Theorem G [310] in our arsenal, we prove one of the most surprising theorems about matrices.

Theorem RMRT
Rank of a Matrix is the Rank of the Transpose
Suppose A is an $m \times n$ matrix. Then $r\left(A\right) = r\left(A^t\right)$. □

Proof We will need a bit of notation before we really get rolling. Write the columns of A as individual vectors, $\mathbf{A}_j$, $1 \leq j \leq n$, and write the rows of A as individual column vectors $\mathbf{R}_i$, $1 \leq i \leq m$. Pictorially then,

$$A = [\mathbf{A}_1|\mathbf{A}_2|\mathbf{A}_3|\dots|\mathbf{A}_n] \qquad\qquad A^t = [\mathbf{R}_1|\mathbf{R}_2|\mathbf{R}_3|\dots|\mathbf{R}_m]$$

Let d denote the rank of A, i.e. $d = r\left(A\right) = \dim\left(\mathcal{R}(A)\right)$. Let $C = \{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3,\, \dots,\, \mathbf{v}_d\} \subseteq \mathbb{C}^m$ be a basis for $\mathcal{R}(A)$.

Every column of A is an element of $\mathcal{R}(A)$ and C is a spanning set for $\mathcal{R}(A)$, so there must be scalars, b_{ij}, $1 \leq i \leq d$, $1 \leq j \leq n$ such that

$$\mathbf{A}_j = b_{1j}\mathbf{v}_1 + b_{2j}\mathbf{v}_2 + b_{3j}\mathbf{v}_3 + \dots + b_{dj}\mathbf{v}_d$$

Define V to be the $m \times d$ matrix whose columns are the vectors $\mathbf{v}_i$, $1 \leq i \leq d$. Let B be the $d \times n$ matrix whose entries are the scalars, b_{ij}. More precisely, $[B]_{ij} = b_{ij}$, $1 \leq i \leq d$, $1 \leq j \leq n$. Then the previous equation expresses column j of A as a linear combination of the columns of V, where the scalars come from column j of B. Let $\mathbf{B}_j$ denote column j of B and then we are in a position to use Definition MVP [195] to write

$$\mathbf{A}_j = V\mathbf{B}_j \qquad\qquad 1 \leq j \leq n$$

Since column j of A is the product of the matrix V with column j of B, Definition MM [197] tells us that $A = VB$. We are now in position to do all of the hard work in this proof — take the transpose of both sides of this equation.

$$\begin{aligned} A &= VB \\ A^t &= (VB)^t \\ [\mathbf{R}_1|\mathbf{R}_2|\mathbf{R}_3|\dots|\mathbf{R}_m] &= B^tV^t && \text{Theorem MMT [205]} \end{aligned}$$

Column i of A^t is row i of A, and by Definition MM [197] this row (expressed as a column vector) is equal to the product of the $n \times d$ matrix B^t with column i of V^t. Applying Definition MVP [195] we see that then each row of A is a linear combination of the d columns of B^t. This is the key observation in this proof.

Since each row of A is a linear combination of the d columns of B^t, and the rows of A are a spanning set for the row space of A, $\nabla s(A)$, we can conclude that the d columns of B^t are a spanning set for $\nabla s(A)$. Now we need Theorem G [310] (or Theorem SSLD [293]). Any basis for the row space of A cannot have more vectors than the size of a spanning set (or it would be linearly dependent), so the dimension of the row space cannot exceed d. Using Definition RSM [183],

$$r\left(A^t\right) = \dim\left(\mathcal{R}\left(A^t\right)\right) = \dim\left(\nabla s(A)\right) \leq d = r\left(A\right)$$

so $r\left(A^t\right) \leq r\left(A\right)$.

This relationship only assumed that A is a matrix. It would apply equally well to A^t, in which case we use Theorem TT [165] and write

$$r\left(A\right) = r\left(\left(A^t\right)^t\right) \leq r\left(A^t\right).$$

These two inequalities together imply that $r\left(A\right) = r\left(A^t\right)$. ■

This says that the row space and the column space of a matrix have the same dimension, which may be 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 [313], 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 [510] has a 4×7 coefficient matrix which row-reduces to

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

so the rank is 3. Row-reducing the transpose yields

$$\begin{bmatrix} \boxed{1} & 0 & 0 & -\frac{31}{7} \\ 0 & \boxed{1} & 0 & \frac{12}{7} \\ 0 & 0 & \boxed{1} & \frac{13}{7} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

demonstrating that the rank of the transpose is also 3. ⊚

Subsection OBC
Orthonormal Bases and Coordinates

We learned about orthogonal sets of vectors in $\mathbb{C}^m$ back in Section O [145], and we also learned that they were automatically linearly independent (Theorem OSLI [153]). When they also span a subspace of $\mathbb{C}^m$, then they are bases. And when the set is orthonormal, then they are incredibly nice bases. 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 GSPCV [153]) and obtain a linearly independent set T such that $\mathcal{S}p(T) = \mathcal{S}p(B) = W$ 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 [155], followed by Example ONTV [156], illustrates this process.

Orthogonal matrices (Definition OM [229]) are another good source of orthonormal bases (and vice versa). Suppose that Q is an orthogonal matrix of size n. Then the n columns of Q form an orthonormal set (Theorem COMOS [231]) that is therefore linearly independent (Theorem OSLI [153]). Since Q is invertible (Theorem OMI [230]), we know Q is nonsingular (Theorem NSI [228]), and then the columns of Q span $\mathbb{C}^n$ (Theorem RNSM [180]). So the columns of an orthogonal matrix of size n are an orthonormal basis for $\mathbb{C}^n$.

Why all the fuss about orthonormal bases? Theorem VRRB [287] 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 $B = \{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3,\, \ldots,\, \mathbf{v}_p\}$ is an orthonormal basis of the subspace W of $\mathbb{C}^m$. For any $\mathbf{w} \in W$,

$$\mathbf{w} = \langle \mathbf{w},\, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{w},\, \mathbf{v}_2 \rangle \mathbf{v}_2 + \langle \mathbf{w},\, \mathbf{v}_3 \rangle \mathbf{v}_3 + \cdots + \langle \mathbf{w},\, \mathbf{v}_p \rangle \mathbf{v}_p$$ □

Proof Because B is a basis of W, Theorem VRRB [287] tells us that we can write $\mathbf{w}$ 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 $\mathbf{w}$ 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,

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_p\mathbf{v}_p$$

and compute,

$$\begin{aligned}
\langle \mathbf{w},\, \mathbf{v}_i \rangle &= \left\langle \sum_{k=1}^{p} a_k\mathbf{v}_k,\, \mathbf{v}_i \right\rangle \\
&= \sum_{k=1}^{p} \langle a_k\mathbf{v}_k,\, \mathbf{v}_i \rangle && \text{Theorem IPVA [147]} \\
&= \sum_{k=1}^{p} a_k \langle \mathbf{v}_k,\, \mathbf{v}_i \rangle && \text{Theorem IPSM [148]} \\
&= a_i \langle \mathbf{v}_i,\, \mathbf{v}_i \rangle + \sum_{k\neq i} a_k \langle \mathbf{v}_k,\, \mathbf{v}_i \rangle && \text{Isolate term with } k = i \\
&= a_i(1) + \sum_{k\neq i} a_k(0) && T \text{ orthonormal} \\
&= a_i
\end{aligned}$$

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

$$\{\mathbf{x}_1,\, \mathbf{x}_2,\, \mathbf{x}_3,\, \mathbf{x}_4\} = \left\{ \begin{bmatrix} 1+i \\ 1 \\ 1-i \\ i \end{bmatrix},\, \begin{bmatrix} 1+5i \\ 6+5i \\ -7-i \\ 1-6i \end{bmatrix},\, \begin{bmatrix} -7+34i \\ -8-23i \\ -10+22i \\ 30+13i \end{bmatrix},\, \begin{bmatrix} -2-4i \\ 6+i \\ 4+3i \\ 6-i \end{bmatrix} \right\}$$

was proposed, and partially verified, as an orthogonal set in Example AOS [152]. Let's scale each vector to norm 1, so as to form an orthonormal basis of $\mathbb{C}^4$. (Notice that by Theorem OSLI [153] the set is linearly independent. Since we know the dimension of $\mathbb{C}^4$

is 4, Theorem G [310] tells us the set is just the right size to be a basis of $\mathbb{C}^4$.) The norms of these vectors are,

$$\|\mathbf{x}_1\| = \sqrt{6} \qquad \|\mathbf{x}_2\| = \sqrt{174} \qquad \|\mathbf{x}_3\| = \sqrt{3451} \qquad \|\mathbf{x}_4\| = \sqrt{119}$$

So an orthonormal basis is

$$\begin{aligned} B &= \{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\mathbf{v}_4\} \\ &= \left\{ \frac{1}{\sqrt{6}}\begin{bmatrix}1+i\\1\\1-i\\i\end{bmatrix},\ \frac{1}{\sqrt{174}}\begin{bmatrix}1+5i\\6+5i\\-7-i\\1-6i\end{bmatrix},\ \frac{1}{\sqrt{3451}}\begin{bmatrix}-7+34i\\-8-23i\\-10+22i\\30+13i\end{bmatrix},\ \frac{1}{\sqrt{119}}\begin{bmatrix}-2-4i\\6+i\\4+3i\\6-i\end{bmatrix} \right\} \end{aligned}$$

Now, choose any vector from $\mathbb{C}^4$, say $\mathbf{w} = \begin{bmatrix}2\\-3\\1\\4\end{bmatrix}$, and 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 [315] guarantees that

$$\begin{aligned} \begin{bmatrix}2\\-3\\1\\4\end{bmatrix} &= \frac{-5i}{\sqrt{6}}\left(\frac{1}{\sqrt{6}}\begin{bmatrix}1+i\\1\\1-i\\i\end{bmatrix}\right) + \frac{-19+30i}{\sqrt{174}}\left(\frac{1}{\sqrt{174}}\begin{bmatrix}1+5i\\6+5i\\-7-i\\1-6i\end{bmatrix}\right) \\ &\quad + \frac{120-211i}{\sqrt{3451}}\left(\frac{1}{\sqrt{3451}}\begin{bmatrix}-7+34i\\-8-23i\\-10+22i\\30+13i\end{bmatrix}\right) + \frac{6+12i}{\sqrt{119}}\left(\frac{1}{\sqrt{119}}\begin{bmatrix}-2-4i\\6+i\\4+3i\\6-i\end{bmatrix}\right) \end{aligned}$$

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

$$\{\mathbf{x}_1,\,\mathbf{x}_2,\,\mathbf{x}_3\} = \left\{ \begin{bmatrix}1\\2\\1\end{bmatrix},\ \begin{bmatrix}-1\\0\\1\end{bmatrix},\ \begin{bmatrix}2\\1\\1\end{bmatrix} \right\}$$

is a linearly independent set, which the Gram-Schmidt Process (Theorem GSPCV [153]) converts to an orthogonal set, and which can then be converted to the orthonormal set,

$$\{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3\} = \left\{ \frac{1}{\sqrt{6}}\begin{bmatrix}1\\2\\1\end{bmatrix},\ \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\0\\1\end{bmatrix},\ \frac{1}{\sqrt{3}}\begin{bmatrix}1\\-1\\1\end{bmatrix} \right\}$$

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 [146]) 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. Its Theorem COB [315] that tells us how to do this.

Suppose that we choose $\mathbf{w} = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}$. Compute

$$\langle \mathbf{w},\, \mathbf{v}_1 \rangle = \frac{5}{\sqrt{6}} \qquad \langle \mathbf{w},\, \mathbf{v}_2 \rangle = \frac{3}{\sqrt{2}} \qquad \langle \mathbf{w},\, \mathbf{v}_3 \rangle = \frac{8}{\sqrt{3}}$$

then Theorem COB [315] 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. ⊚

Subsection READ Reading Questions

1. Why does Theorem G [310] have the title it does?
2. What is so surprising about Theorem RMRT [313]?
3. Why is an orthonormal basis desirable?

Subsection EXC
Exercises

T60 Contributed by Joe Riegsecker
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. Solution [321]

TODO: Every subspace has a finite basis
TODO: Theorem ELIS "forever" in a finite-dim vsp?
TODO: rank of mxn matrix less than min of m and n
TODO: LS(A,b) consistent iff rank A = rank A—b

Subsection SOL
Solutions

T60 Exercise [319] Contributed by Robert Beezer
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 [310] 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 $\mathbf{w} \in U \cap V$.

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.

D: Determinants

Section DM
Determinants of Matrices

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. We'll begin with a definition, do some computations and then establish some properties. 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_{ij} is the $(m-1) \times (n-1)$ matrix obtained from A by removing row i and column j. △

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

$$A_{23} = \begin{bmatrix} 1 & -2 & 9 \\ 3 & 5 & 1 \end{bmatrix} \qquad A_{31} \begin{bmatrix} -2 & 3 & 9 \\ -2 & 0 & 1 \end{bmatrix}$$ ⊚

Definition DM
Determinant
Suppose A is a square matrix. Then its **determinant**, $\det(A) = |A|$, is an element of $\mathbb{C}$

defined recursively by:

If $A = \begin{bmatrix} a \end{bmatrix}$ is a 1×1 matrix, then $\det(A) = a$.

If $A = (a_{ij})$ is a matrix of size n with $n \geq 2$, then

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + a_{13} \det(A_{13}) - \cdots + (-1)^{n+1} a_{1n} \det(A_{1n}) \qquad \triangle$$

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 $10! = 10\times 9\times 8\times 7\times 6\times 5\times 4\times 3\times 2 = 3,628,800$ 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.

Lets 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

$$\begin{aligned}
\det(A) = |A| &= \begin{vmatrix} 3 & 2 & -1 \\ 4 & 1 & 6 \\ -3 & -1 & 2 \end{vmatrix} \\
&= 3 \begin{vmatrix} 1 & 6 \\ -1 & 2 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ -3 & 2 \end{vmatrix} + (-1) \begin{vmatrix} 4 & 1 \\ -3 & -1 \end{vmatrix} \\
&= 3\left(1\left|2\right| - 6\left|-1\right|\right) - 2\left(4\left|2\right| - 6\left|-3\right|\right) - \left(4\left|-1\right| - 1\left|-3\right|\right) \\
&= 3\left(1(2) - 6(-1)\right) - 2\left(4(2) - 6(-3)\right) - \left(4(-1) - 1(-3)\right) \\
&= 24 - 52 + 1 \\
&= -27
\end{aligned}$$

⊚

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 [323],

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a\left|d\right| - b\left|c\right| = ad - bc \qquad \blacksquare$$

Do you recall seeing the expression $ad - bc$ before? (Hint: Theorem TTMI [218])

Subsection CD
Computing Determinants

For any given matrix, there are a variety of ways to compute the determinant, by "expanding" about any row or column. The determinants of the submatrices used in these computations are used so often that they have their own names. The first is the determinant of a submatrix, the second differs only by a sign.

Definition MIM
Minor In a Matrix
Suppose A is an $n \times n$ matrix and A_{ij} is the $(n-1) \times (n-1)$ submatrix formed by removing row i and column j. Then the **minor** for A at location $i\,j$ is the determinant of the submatrix, $M_{A,ij} = \det(A_{ij})$. △

Definition CIM
Cofactor In a Matrix
Suppose A is an $n \times n$ matrix and A_{ij} is the $(n-1) \times (n-1)$ submatrix formed by removing row i and column j. Then the **cofactor** for A at location $i\,j$ is the signed determinant of the submatrix, $C_{A,ij} = (-1)^{i+j}\det(A_{ij})$. △

Example MC
Minors and cofactors
For the matrix,

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

we have minors

$$M_{A,4,2} = \begin{vmatrix} 2 & 2 & 1 \\ -1 & 3 & -1 \\ 3 & 0 & 5 \end{vmatrix} = 2(15) - 2(-2) + 1(-9) = 25$$

$$M_{A,3,4} = \begin{vmatrix} 2 & 4 & 2 \\ -1 & 2 & 3 \\ 3 & 6 & 3 \end{vmatrix} = 2(-12) - 4(-12) + 2(-12) = 0$$

and so two cofactors are

$$C_{A,4,2} = (-1)^{4+2} M_{A,4,2} = (1)(25) = 25$$
$$C_{A,3,4} = (-1)^{3+4} M_{A,3,4} = (-1)(0) = 0$$

A third cofactor is

$$C_{A,1,2} = (-1)^{1+2} M_{A,1,2} = (-1)\begin{vmatrix} -1 & 3 & -1 \\ 3 & 0 & 5 \\ 3 & 3 & 2 \end{vmatrix}$$
$$= (-1)((-1)(-15) - (3)(-9) + (-1)(9)) = -33$$

⊚

With this notation in hand, we can state

Theorem DERC
Determinant Expansion about Rows and Columns
Suppose that $A = (a_{ij})$ is a square matrix of size n. Then

$$\det(A) = a_{i1}C_{A,i1} + a_{i2}C_{A,i2} + a_{i3}C_{A,i3} + \cdots + a_{in}C_{A,in} \qquad 1 \le i \le n$$

which is known as **expansion** about row i, and

$$\det(A) = a_{1j}C_{A,1j} + a_{2j}C_{A,2j} + a_{3j}C_{A,3j} + \cdots + a_{nj}C_{A,nj} \qquad 1 \le j \le n$$

which is known as **expansion** about column j. □

Proof TODO ■

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 DERC [326] with $i = 4$) yields,

$$\begin{aligned} |A| &= (4)(-1)^{4+1}\begin{vmatrix} 3 & 0 & 1 \\ -2 & 0 & 1 \\ 3 & -2 & -1 \end{vmatrix} + (1)(-1)^{4+2}\begin{vmatrix} -2 & 0 & 1 \\ 9 & 0 & 1 \\ 1 & -2 & -1 \end{vmatrix} \\ &\quad + (2)(-1)^{4+3}\begin{vmatrix} -2 & 3 & 1 \\ 9 & -2 & 1 \\ 1 & 3 & -1 \end{vmatrix} + (6)(-1)^{4+4}\begin{vmatrix} -2 & 3 & 0 \\ 9 & -2 & 0 \\ 1 & 3 & -2 \end{vmatrix} \\ &= (-4)(10) + (1)(-22) + (-2)(61) + 6(46) = 92 \end{aligned}$$

while expanding about column 3 (Theorem DERC [326] with $j = 3$) gives

$$\begin{aligned} |A| &= (0)(-1)^{1+3}\begin{vmatrix} 9 & -2 & 1 \\ 1 & 3 & -1 \\ 4 & 1 & 6 \end{vmatrix} + (0)(-1)^{2+3}\begin{vmatrix} -2 & 3 & 1 \\ 1 & 3 & -1 \\ 4 & 1 & 6 \end{vmatrix} + \\ &\quad (-2)(-1)^{3+3}\begin{vmatrix} -2 & 3 & 1 \\ 9 & -2 & 1 \\ 1 & 3 & -1 \end{vmatrix} + (2)(-1)^{4+3}\begin{vmatrix} -2 & 3 & 1 \\ 9 & -2 & 1 \\ 4 & 1 & 6 \end{vmatrix} \\ &= 0 + 0 + (-2)(61) + (-2)(-107) = 92 \end{aligned}$$

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.

$$\begin{aligned}
\det(T) &= \begin{vmatrix} 2 & 3 & -1 & 3 & 3 \\ 0 & -1 & 5 & 2 & -1 \\ 0 & 0 & 3 & 9 & 2 \\ 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 5 \end{vmatrix} \\
&= 2(-1)^{1+1} \begin{vmatrix} -1 & 5 & 2 & -1 \\ 0 & 3 & 9 & 2 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{vmatrix} \\
&= 2(-1)(-1)^{1+1} \begin{vmatrix} 3 & 9 & 2 \\ 0 & -1 & 3 \\ 0 & 0 & 5 \end{vmatrix} \\
&= 2(-1)(3)(-1)^{1+1} \begin{vmatrix} -1 & 3 \\ 0 & 5 \end{vmatrix} \\
&= 2(-1)(3)(-1)(-1)^{1+1} \begin{vmatrix} 5 \end{vmatrix} \\
&= 2(-1)(3)(-1)(5) = 30
\end{aligned}$$

⊚

Subsection PD
Properties of Determinants

The determinant is of some interest by itself, but it is of the greatest use when employed to *determine* properties of matrices. To that end, we list some theorems here. Unfortunately, mostly without proof at the moment.

Theorem DT
Determinant of the Transpose
Suppose that A is a square matrix. Then $\det(A^t) = \det(A)$. □

Proof We will prove this result by induction on the size of the matrix. For a matrix of size 1, the transpose and the matrix itself are equal, so matter what the definition of a determinant might be, their determinants are equal.

Now suppose the theorem is true for matrices of size $n-1$. By Theorem DERC [326] we can write the determinant as a product of entries from the first row with their cofactors and then sum these products. These cofactors are signed determinants of matrices of size $n-1$, which by the induction hypothesis, are equal to the determinant of their transposes, and commutativity in the sum in the exponent of -1 means the cofactor is equal to a cofactor of the transpose.

$$\begin{aligned}
\det(A) &= [A]_{11} C_{A,11} + [A]_{12} C_{A,12} \\
&\quad + [A]_{13} C_{A,13} + \cdots + [A]_{1n} C_{A,1n} && \text{Theorem DERC [326], row 1} \\
&= \left[A^t\right]_{11} C_{A,11} + \left[A^t\right]_{21} C_{A,12} \\
&\quad + \left[A^t\right]_{31} C_{A,13} + \cdots + \left[A^t\right]_{n1} C_{A,1n} && \text{Definition TM [163]} \\
&= \left[A^t\right]_{11} C_{A^t,11} + \left[A^t\right]_{21} C_{A^t,21} \\
&\quad + \left[A^t\right]_{31} C_{A^t,31} + \cdots + \left[A^t\right]_{n1} C_{A^t,n1} && \text{Induction hypothesis} \\
&= \det\left(A^t\right) && \text{Theorem DERC [326], column 1}
\end{aligned}$$
■

Theorem DRMM
Determinant Respects Matrix Multiplication
Suppose that A and B are square matrices of size n. Then $\det(AB) = \det(A)\det(B)$. □

Proof TODO: ■

Its an amazing thing that matrix multiplication and the determinant interact this way. Might it also be true that $\det(A+B) = \det(A) + \det(B)$?

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 TODO: ■

For the case of 2×2 matrices you might compare the application of Theorem SMZD [328] with the combination of the results stated in Theorem DMST [324] and Theorem TTMI [218].

Example ZNDAB
Zero and nonzero determinant, Archetypes A and B
The coefficient matrix in Archetype A [473] has a zero determinant (check this!) while the coefficient matrix Archetype B [478] has a nonzero determinant. These matrices are singular and nonsingular, respectively. This is exactly what Theorem SMZD [328] says, and continues our list of contrasts between these two archetypes. ⊚

Since Theorem SMZD [328] is an equivalence (Technique E [54]) we can expand on our growing list of equivalence about nonsingular matrices.

Theorem NSME7
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. The range of A is $\mathbb{C}^n$, $\mathcal{R}(A) = \mathbb{C}^n$.
7. A is invertible.
8. The columns of A are a basis for $\mathbb{C}^n$.
9. The rank of A is n, $r\left(A\right) = n$.
10. The nullity of A is zero, $n\left(A\right) = 0$.
11. The determinant of A is nonzero, $\det\left(A\right) \neq 0$. □

Proof Theorem SMZD [328] says A is singular if and only if $\det\left(A\right) = 0$. 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 $\det\left(A\right) \neq 0$. This allows us to add a new statement to the list. ■

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 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 very quickly becomes so excessive as to be impractical.

Subsection READ
Reading Questions

1. Compute the determinant of the matrix

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

2. What is our latest addition to the NSMExx series of theorems?

3. What is amazing about the interaction between matrix multiplication and the determinant?

E: Eigenvalues

Section EE
Eigenvalues and Eigenvectors

When we have a square matrix of size n, A, and we multiply it by a vector from $\mathbb{C}^n$, $\mathbf{x}$, to form the matrix-vector product (Definition MVP [195]), 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.

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 from $\mathbb{C}^n$, and λ is a scalar from $\mathbb{C}$ such that

$$A\mathbf{x} = \lambda\mathbf{x}$$

Then we say $\mathbf{x}$ is an **eigenvector** of A with **eigenvalue** λ. △

Before going any further, perhaps we should convince you that such things ever happen at all. Believe 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

$$\mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 5 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} -3 \\ 4 \\ -10 \\ 4 \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} -3 \\ 7 \\ 0 \\ 8 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 1 \\ -1 \\ 4 \\ 0 \end{bmatrix}$$

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 $\mathbf{x}$ 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 $\mathbf{y}$ 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 $\mathbf{z}$ 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 $\mathbf{w}$ 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

$$\begin{aligned}
A\mathbf{u} &= A(30\mathbf{x}) && \text{Substitution}\\
&= 30A\mathbf{x} && \text{Distributivity}\\
&= 30(4\mathbf{x}) && \mathbf{x}\text{ is an eigenvector of }A\\
&= 4(30\mathbf{x}) && \text{Associativity, Commutativity}\\
&= 4\mathbf{u} && \text{Substitution}
\end{aligned}$$

so that $\mathbf{u}$ 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,

$$\begin{aligned}
A\mathbf{v} &= A(\mathbf{z} + \mathbf{w}) && \text{Substitution}\\
&= A\mathbf{z} + A\mathbf{w} && \text{Distributivity}\\
&= 2\mathbf{z} + 2\mathbf{w} && \mathbf{z},\ \mathbf{w}\text{ eigenvectors of }A\\
&= 2(\mathbf{z} + \mathbf{w}) && \text{Distributivity}\\
&= 2\mathbf{v} && \text{Substitution}
\end{aligned}$$

so that $\mathbf{v}$ 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 $\lambda = 0$, which essentially is a result of the fact that $A\mathbf{y} = \mathbf{0}$. But this also means that $\mathbf{y} \in \mathcal{N}(A)$. There would appear to be a connection here also. ⊚

Example SEE [332] 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 [353], but for now 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 in a polynomial. So it is with matrices. We can add and subtract, we can multiply by scalars, and we can form powers by repeated uses 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 of a polynomial will preserve the size of the 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.

$$\begin{aligned}
D^0 = I_3 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
D^1 = D &= \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \\
D^2 = DD^1 &= \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -1 & -6 \\ 5 & 1 & 0 \\ 1 & -8 & -7 \end{bmatrix} \\
D^3 = DD^2 &= \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & -1 & -6 \\ 5 & 1 & 0 \\ 1 & -8 & -7 \end{bmatrix} = \begin{bmatrix} 19 & -12 & -8 \\ -4 & 15 & 8 \\ 12 & -4 & 11 \end{bmatrix} \\
D^4 = DD^3 &= \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 19 & -12 & -8 \\ -4 & 15 & 8 \\ 12 & -4 & 11 \end{bmatrix} = \begin{bmatrix} -7 & 49 & 54 \\ -5 & -4 & -30 \\ -49 & 47 & 43 \end{bmatrix}
\end{aligned}$$

Then

$$\begin{aligned}
p(D) &= 14 + 19D - 3D^2 - 7D^3 + D^4 \\
&= 14 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + 19 \begin{bmatrix} -1 & 3 & 2 \\ 1 & 0 & -2 \\ -3 & 1 & 1 \end{bmatrix} - 3 \begin{bmatrix} -2 & -1 & -6 \\ 5 & 1 & 0 \\ 1 & -8 & -7 \end{bmatrix} \\
&\quad - 7 \begin{bmatrix} 19 & -12 & -8 \\ -4 & 15 & 8 \\ 12 & -4 & 11 \end{bmatrix} + \begin{bmatrix} -7 & 49 & 54 \\ -5 & -4 & -30 \\ -49 & 47 & 43 \end{bmatrix} \\
&= \begin{bmatrix} -139 & 193 & 166 \\ 27 & -98 & -124 \\ -193 & 118 & 20 \end{bmatrix}
\end{aligned}$$

Notice that $p(x)$ factors as

$$p(x) = 14 + 19x - 3x^2 - 7x^3 + x^4 = (x-2)(x-7)(x+1)^2$$

Because A commutes with itself ($AA = AA$), we can use distributivity of matrix multiplication across matrix addition without being careful with any of the matrix products, and just as easily evaluate $p(D)$ using the factored form of $p(x)$,

$$
\begin{aligned}
p(x) &= 14 + 19D - 3D^2 - 7D^3 + D^4 = (D - 2I_3)(D - 7I_3)(D + I_3)^2 \\
&= \begin{bmatrix} -3 & 3 & 2 \\ 1 & -2 & -2 \\ -3 & 1 & -1 \end{bmatrix} \begin{bmatrix} -8 & 3 & 2 \\ 1 & -7 & -2 \\ -3 & 1 & -6 \end{bmatrix} \begin{bmatrix} 0 & 3 & 2 \\ 1 & 1 & -2 \\ -3 & 1 & 2 \end{bmatrix}^2 \\
&= \begin{bmatrix} -139 & 193 & 166 \\ 27 & -98 & -124 \\ -193 & 118 & 20 \end{bmatrix}
\end{aligned}
$$

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. ⊚

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 [362], we will determine the maximum number of eigenvalues a matrix may have.

The determinant (Definition D [293]) will be a powerful tool in Subsection EE.CEE [339] 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 $\mathbf{x}$ as *any* nonzero vector from $\mathbb{C}^n$. (Notice how much latitude we have in our choice of $\mathbf{x}$. Only the zero vector is off-limits.) Consider the set

$$S = \left\{\mathbf{x},\, A\mathbf{x},\, A^2\mathbf{x},\, A^3\mathbf{x},\, \ldots,\, A^n\mathbf{x}\right\}$$

This is a set of $n + 1$ vectors from $\mathbb{C}^n$, so by Theorem MVSLD [133], 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,

$$a_0\mathbf{x} + a_1A\mathbf{x} + a_2A^2\mathbf{x} + a_3A^3\mathbf{x} + \cdots + a_nA^n\mathbf{x} = \mathbf{0}$$

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 [249], 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

$$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_mx^m$$

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

$$\begin{aligned}
\mathbf{0} &= a_0\mathbf{x} + a_1A\mathbf{x} + a_2A^2\mathbf{x} + a_3A^3\mathbf{x} + \cdots + a_nA^n\mathbf{x} \\
&= a_0\mathbf{x} + a_1A\mathbf{x} + a_2A^2\mathbf{x} + a_3A^3\mathbf{x} + \cdots + a_mA^m\mathbf{x} && a_i = 0 \text{ for } i > m \\
&= \left(a_0I_n + a_1A + a_2A^2 + a_3A^3 + \cdots + a_mA^m\right)\mathbf{x} && \text{Theorem MMDAA [201]} \\
&= p(A)\mathbf{x} && \text{Definition of } p(x) \\
&= (A - b_mI_n)(A - b_{m-1}I_n)\cdots(A - b_3I_n)(A - b_2I_n)(A - b_1I_n)\mathbf{x}
\end{aligned}$$

Let k be the smallest integer such that

$$(A - b_kI_n)(A - b_{k-1}I_n)\cdots(A - b_3I_n)(A - b_2I_n)(A - b_1I_n)\mathbf{x} = \mathbf{0}.$$

From the preceding equation, we know that $k \leq m$. Define the vector $\mathbf{z}$ 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 event that $k = 1$, we take $\mathbf{z} = \mathbf{x}$, and $\mathbf{z}$ is still nonzero. Now

$$(A - b_kI_n)\mathbf{z} = (A - b_kI_n)(A - b_{k-1}I_n)\cdots(A - b_3I_n)(A - b_2I_n)(A - b_1I_n)\mathbf{x} = \mathbf{0}$$

which we can rearrange as

$$\begin{aligned}
(A - b_kI_n)\mathbf{z} &= \mathbf{0} \\
A\mathbf{z} - b_kI_n\mathbf{z} &= \mathbf{0} && \text{Theorem MMDAA [201]} \\
A\mathbf{z} - b_k\mathbf{z} &= \mathbf{0} && \text{Theorem MMIM [201]} \\
A\mathbf{z} &= b_k\mathbf{z}
\end{aligned}$$

Since $\mathbf{z} \neq \mathbf{0}$, this final line says that $\mathbf{z}$ is an eigenvector of A for the eigenvalue $\lambda = b_k$, so A does have at least one eigenvalue. ■

The proof of Theorem EMHE [335] 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 [335], 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

$$\begin{aligned} S &= \left\{\mathbf{x},\, A\mathbf{x},\, A^2\mathbf{x},\, A^3\mathbf{x},\, A^4\mathbf{x},\, A^5\mathbf{x}\right\} \\ &= \left\{ \begin{bmatrix} 3 \\ 0 \\ 3 \\ -5 \\ 4 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \\ -4 \\ 4 \\ -6 \end{bmatrix}, \begin{bmatrix} 6 \\ -6 \\ 6 \\ -2 \\ 10 \end{bmatrix}, \begin{bmatrix} -10 \\ 14 \\ -10 \\ -2 \\ -18 \end{bmatrix}, \begin{bmatrix} 18 \\ -30 \\ 18 \\ 10 \\ 34 \end{bmatrix}, \begin{bmatrix} -34 \\ 62 \\ -34 \\ -26 \\ -66 \end{bmatrix} \right\} \end{aligned}$$

is guaranteed to be linearly dependent, as it has six vectors from $\mathbb{C}^5$ (Theorem MVSLD [133]). 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,

$$\begin{bmatrix} 3 & -4 & 6 & -10 & 18 & -34 \\ 0 & 2 & -6 & 14 & -30 & 62 \\ 3 & -4 & 6 & -10 & 18 & -34 \\ -5 & 4 & -2 & -2 & 10 & -26 \\ 4 & -6 & 10 & -18 & 34 & -66 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & -2 & 6 & -14 & 30 \\ 0 & \boxed{1} & -3 & 7 & -15 & 31 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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 [335] 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

$$\begin{aligned} \mathbf{0} &= 16\mathbf{x} + 12A\mathbf{x} - 8A^2\mathbf{x} - 3A^3\mathbf{x} + A^4\mathbf{x} + 0A^5\mathbf{x} \\ \mathbf{0} &= \left(16 + 12A - 8A^2 - 3A^3 + A^4\right)\mathbf{x} \end{aligned}$$

So we define $p(x) = 16 + 12x - 8x^2 - 3x^3 + x^4$, and as advertised in the proof of Theorem EMHE [335], 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 [335],

$$(A + 1I_5)\mathbf{x} = \begin{bmatrix} -6 & -1 & 11 & 0 & -4 \\ 4 & 2 & 0 & 2 & 0 \\ -10 & -1 & 15 & 0 & -4 \\ 8 & 2 & -15 & 0 & 5 \\ -10 & -1 & 16 & 0 & -5 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \\ 3 \\ -5 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \\ -1 \\ -2 \end{bmatrix}$$

$$(A - 2I_5)(A + 1I_5)\mathbf{x} = \begin{bmatrix} -9 & -1 & 11 & 0 & -4 \\ 4 & -1 & 0 & 2 & 0 \\ -10 & -1 & 12 & 0 & -4 \\ 8 & 2 & -15 & -3 & 5 \\ -10 & -1 & 16 & 0 & -8 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ -1 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 4 \\ -8 \\ 4 \\ 4 \\ 8 \end{bmatrix}$$

$$(A + 2I_5)(A - 2I_5)(A + 1I_5)\mathbf{x} = \begin{bmatrix} -5 & -1 & 11 & 0 & -4 \\ 4 & 3 & 0 & 2 & 0 \\ -10 & -1 & 16 & 0 & -4 \\ 8 & 2 & -15 & 1 & 5 \\ -10 & -1 & 16 & 0 & -4 \end{bmatrix} \begin{bmatrix} 4 \\ -8 \\ 4 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

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 $A\mathbf{z}$. If you work through this example with your own choice of the vector $\mathbf{x}$ (strongly recommended) then the eigenvalue you will find may be different, but will be in the set $\{3,\ 0,\ 1,\ -1,\ -2\}$. ⊚

Subsection CEE
Computing Eigenvalues and Eigenvectors

Fortunately, we need not rely on the procedure of Theorem EMHE [335] each time we need an eigenvalue. It is the determinant, and specifically Theorem SMZD [328], that provide the main tool. First a key definition.

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

$$p_A(x) = \det(A - xI_n)$$

△

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

$$\begin{aligned}
p_F(x) &= \det(F - xI_3) \\
&= \begin{vmatrix} -13-x & -8 & -4 \\ 12 & 7-x & 4 \\ 24 & 16 & 7-x \end{vmatrix} && \text{Definition CP [339]} \\
&= (-13-x)\begin{vmatrix} 7-x & 4 \\ 16 & 7-x \end{vmatrix} + (-8)(-1)\begin{vmatrix} 12 & 4 \\ 24 & 7-x \end{vmatrix} \\
&\quad + (-4)\begin{vmatrix} 12 & 7-x \\ 24 & 16 \end{vmatrix} && \text{Definition DM [323]} \\
&= (-13-x)((7-x)(7-x) - 4(16)) \\
&\quad + (-8)(-1)(12(7-x) - 4(24)) \\
&\quad + (-4)(12(16) - (7-x)(24)) && \text{Theorem DMST [324]} \\
&= 3 + 5x + x^2 - x^3 \\
&= -(x-3)(x+1)^2
\end{aligned}$$

⊚

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.

$$\begin{aligned}
p_A(\lambda) = 0 &\iff \det(A - \lambda I_n) = 0 && \text{Definition CP [339]} \\
&\iff A - \lambda I_n \text{ is singular} && \text{Theorem SMZD [328]} \\
&\iff \text{there exists } \mathbf{x} \neq \mathbf{0} \text{ so that } (A - \lambda I_n)\mathbf{x} = \mathbf{0} && \text{Definition NM [75]} \\
&\iff A\mathbf{x} - \lambda I_n\mathbf{x} = \mathbf{0} && \text{Theorem MMDAA [201]} \\
&\iff A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} && \text{Theorem MMIM [201]} \\
&\iff A\mathbf{x} = \lambda\mathbf{x},\ \mathbf{x} \neq \mathbf{0} \\
&\iff \lambda \text{ is an eigenvalue of } A && \text{Definition EEM [331]}
\end{aligned}$$

■

Example EMS3
Eigenvalues of a matrix, size 3

In Example CPMS3 [340] 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 [340], $\lambda = 3$ 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, $E_A(\lambda)$, is the set of all the eigenvectors of A for λ, with the addition of the zero vector. △

Example SEE [332] 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 $E_A(\lambda)$ is a subspace of the vector space $\mathbb{C}^n$. □

Proof We will check the three conditions of Theorem TSS [255]. First, Definition EM [341] explicitly includes the zero vector in $E_A(\lambda)$, so the set is non-empty.

Suppose that $\mathbf{x}, \mathbf{y} \in E_A(\lambda)$, that is, $\mathbf{x}$ and $\mathbf{y}$ are two eigenvectors of A for λ. Then

$$\begin{aligned} A(\mathbf{x}+\mathbf{y}) &= A\mathbf{x} + A\mathbf{y} && \text{Theorem MMDAA [201]} \\ &= \lambda\mathbf{x} + \lambda\mathbf{y} && \mathbf{x},\ \mathbf{y} \text{ eigenvectors of } A \\ &= \lambda(\mathbf{x}+\mathbf{y}) && \text{Distributivity} \end{aligned}$$

So this says either $\mathbf{x}+\mathbf{y} = \mathbf{0}$, or $\mathbf{x}+\mathbf{y}$ is an eigenvector of A for λ. So, in either event, $\mathbf{x}+\mathbf{y} \in E_A(\lambda)$, and we have additive closure.

Suppose that $\alpha \in \mathbb{C}$, and that $\mathbf{x} \in E_A(\lambda)$, that is, $\mathbf{x}$ is an eigenvector of A for λ. Then

$$\begin{aligned} A(\alpha\mathbf{x}) &= \alpha(A\mathbf{x}) && \text{Theorem MMSMM [202]} \\ &= \alpha\lambda\mathbf{x} && \mathbf{x} \text{ an eigenvector of } A \\ &= \lambda(\alpha\mathbf{x}) && \text{Distributivity} \end{aligned}$$

which says that $\alpha\mathbf{x}$ is an eigenvector of A for λ, so $\alpha\mathbf{x} \in E_A(\lambda)$, and we have scalar closure.

With the three conditions of Theorem TSS [255] met, we know $E_A(\lambda)$ is a subspace. ■

Theorem EMS [341] 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 Technique SE [21]. 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 [341] and $\mathbf{0} \in \mathcal{N}(A - \lambda I_n)$ by Theorem HSC [68]. Now consider a nonzero vector $\mathbf{x} \in \mathbb{C}^n$,

$$\begin{aligned}
\mathbf{x} \in \mathcal{E}_A(\lambda) &\iff A\mathbf{x} = \lambda\mathbf{x} && \text{Definition EM [341]}\\
&\iff A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} &&\\
&\iff A\mathbf{x} - \lambda I_n\mathbf{x} = \mathbf{0} && \text{Theorem MMIM [201]}\\
&\iff (A - \lambda I_n)\,\mathbf{x} = \mathbf{0} && \text{Theorem MMDAA [201]}\\
&\iff \mathbf{x} \in \mathcal{N}(A - \lambda I_n) && \text{Definition NSM [73]}
\end{aligned}$$

■

You might notice the close parallels (and differences) between the proofs of Theorem EMRCP [340] and Theorem EMNS [342]. Since Theorem EMNS [342] describes the set of all the eigenvectors of A as a null space we can use techniques such as Theorem BNS [138] to provide concise descriptions of eigenspaces.

Example ESMS3
Eigenspaces of a matrix, size 3
Example CPMS3 [340] and Example EMS3 [340] 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 the 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 [342]). Theorem BNS [138] then tells us how to write the null space as the span of a basis.

$$\lambda = 3 \qquad F - 3I_3 = \begin{bmatrix} -16 & -8 & -4 \\ 12 & 4 & 4 \\ 24 & 16 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & \frac{1}{2} \\ 0 & \boxed{1} & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

$$E_F(3) = \mathcal{N}(F - 3I_3) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = -1 \qquad F + 1I_3 = \begin{bmatrix} -12 & -8 & -4 \\ 12 & 8 & 4 \\ 24 & 16 & 8 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$E_F(-1) = \mathcal{N}(F + 1I_3) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{2}{3} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{3} \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

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. ⊚

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 [340], Example EMS3 [340] and Example ESMS3 [342]. First, we will sneak in two similar definitions and 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, $p_A(x)$. △

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.

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 $E_A(\lambda)$. △

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,

$$\lambda = 1 \qquad B - 1I_4 = \begin{bmatrix} -3 & 1 & -2 & -4 \\ 12 & 0 & 4 & 9 \\ 6 & 5 & -3 & -4 \\ 3 & -4 & 5 & 9 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & \frac{1}{3} & 0 \\ 0 & \boxed{1} & -1 & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_B(1) = \mathcal{N}(B - 1I_4) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{3} \\ 1 \\ 1 \\ 0 \end{bmatrix}\right\}\right)$$

$$\lambda = 2 \qquad B - 2I_4 = \begin{bmatrix} -4 & 1 & -2 & -4 \\ 12 & -1 & 4 & 9 \\ 6 & 5 & -4 & -4 \\ 3 & -4 & 5 & 8 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 1/2 \\ 0 & \boxed{1} & 0 & -1 \\ 0 & 0 & \boxed{1} & 1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_B(2) = \mathcal{N}(B - 2I_4) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{2} \\ -1 \\ -\frac{1}{2} \\ 1 \end{bmatrix}\right\}\right)$$

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 [374]). ⊚

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,

$$\lambda = 3 \qquad C - 3I_4 = \begin{bmatrix} -2 & 0 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 1 & 1 & -2 & 0 \\ 1 & 1 & 0 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & -1 \\ 0 & \boxed{1} & 0 & -1 \\ 0 & 0 & \boxed{1} & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_C(3) = \mathcal{N}(C - 3I_4) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = 1 \qquad C - 1I_4 = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 1 & 0 & 0 \\ 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_C(1) = \mathcal{N}(C - 1I_4) = \mathcal{S}p\left(\left\{\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix},\ \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = -1 \qquad C + 1I_4 = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 1 \\ 0 & \boxed{1} & 0 & 1 \\ 0 & 0 & \boxed{1} & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_C(-1) = \mathcal{N}(C + 1I_4) = \mathcal{S}p\left(\left\{\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$

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 [362]. ⊚

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

$$p_E\left(x\right) = -16 + 16x + 8x^2 - 16x^3 + 7x^4 - x^5 = -(x-2)^4(x+1)$$

So the eigenvalues are $\lambda = 2,\ -1$ with algebraic multiplicities $\alpha_E\left(2\right) = 4$ and $\alpha_E\left(-1\right) = 1$.

Computing eigenvectors,

$$\lambda = 2 \qquad E - 2I_5 = \begin{bmatrix} 27 & 14 & 2 & 6 & -9 \\ -47 & -24 & -1 & -11 & 13 \\ 19 & 10 & 3 & 4 & -8 \\ -19 & -10 & -3 & -4 & 8 \\ 7 & 4 & 3 & 1 & -5 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 1 & 0 \\ 0 & \boxed{1} & 0 & -\frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & \boxed{1} & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_E\left(2\right) = \mathcal{N}(E - 2I_5) = \mathcal{S}p\left(\left\{ \begin{bmatrix} -1 \\ \frac{3}{2} \\ 0 \\ 1 \\ 0 \end{bmatrix},\ \begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}\right)$$

$$\lambda = -1 \qquad E + 1I_5 = \begin{bmatrix} 30 & 14 & 2 & 6 & -9 \\ -47 & -21 & -1 & -11 & 13 \\ 19 & 10 & 6 & 4 & -8 \\ -19 & -10 & -3 & -1 & 8 \\ 7 & 4 & 3 & 1 & -2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 2 & 0 \\ 0 & \boxed{1} & 0 & -4 & 0 \\ 0 & 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_E\left(-1\right) = \mathcal{N}(E + 1I_5) = \mathcal{S}p\left(\left\{ \begin{bmatrix} -2 \\ 4 \\ -1 \\ 1 \\ 0 \end{bmatrix} \right\}\right)$$

So the eigenspace dimensions yield geometric multiplicities $\gamma_E\left(2\right) = 2$ and $\gamma_E\left(-1\right) = 1$. ⊙

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

$$\begin{aligned} p_F(x) &= -50 + 55x + 13x^2 - 50x^3 + 32x^4 - 9x^5 + x^6 \\ &= (x-2)(x+1)(x^2-4x+5)^2 \\ &= (x-2)(x+1)((x-(2+i))(x-(2-i)))^2 \\ &= (x-2)(x+1)(x-(2+i))^2(x-(2-i))^2 \end{aligned}$$

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,

$\lambda = 2$

$$F - 2I_6 = \begin{bmatrix} -61 & -34 & 41 & 12 & 25 & 30 \\ 1 & 5 & -46 & -36 & -11 & -29 \\ -233 & -119 & 56 & -35 & 75 & 54 \\ 157 & 81 & -43 & 19 & -51 & -39 \\ -91 & -48 & 32 & -5 & 30 & 26 \\ 209 & 107 & -55 & 28 & -69 & -52 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & \boxed{1} & 0 & -\frac{1}{5} \\ 0 & 0 & 0 & 0 & \boxed{1} & \frac{4}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_F(2) = \mathcal{N}(F - 2I_6) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{5} \\ 0 \\ -\frac{3}{5} \\ \frac{1}{5} \\ -\frac{4}{5} \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = -1$$

$$F + 1I_6 = \begin{bmatrix} -58 & -34 & 41 & 12 & 25 & 30 \\ 1 & 8 & -46 & -36 & -11 & -29 \\ -233 & -119 & 59 & -35 & 75 & 54 \\ 157 & 81 & -43 & 22 & -51 & -39 \\ -91 & -48 & 32 & -5 & 33 & 26 \\ 209 & 107 & -55 & 28 & -69 & -49 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & \boxed{1} & 0 & 0 & 0 & -\frac{3}{2} \\ 0 & 0 & \boxed{1} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_F(-1) = \mathcal{N}(F + I_6) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{2} \\ \frac{3}{2} \\ -\frac{1}{2} \\ 0 \\ \frac{1}{2} \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = 2 + i$$

$$F - (2+i)I_6 = \begin{bmatrix} -61-i & -34 & 41 & 12 & 25 & 30 \\ 1 & 5-i & -46 & -36 & -11 & -29 \\ -233 & -119 & 56-i & -35 & 75 & 54 \\ 157 & 81 & -43 & 19-i & -51 & -39 \\ -91 & -48 & 32 & -5 & 30-i & 26 \\ 209 & 107 & -55 & 28 & -69 & -52-i \end{bmatrix}$$

$$\xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & \frac{1}{5}(7+i) \\ 0 & \boxed{1} & 0 & 0 & 0 & \frac{1}{5}(-9-2i) \\ 0 & 0 & \boxed{1} & 0 & 0 & 1 \\ 0 & 0 & 0 & \boxed{1} & 0 & -1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathcal{E}_F(2+i) = \mathcal{N}(F - (2+i)I_6) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{5}(7+i) \\ \frac{1}{5}(9+2i) \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}\right\}\right)$$

$\lambda = 2 - i$

$$F - (2-i)I_6 = \begin{bmatrix} -61+i & -34 & 41 & 12 & 25 & 30 \\ 1 & 5+i & -46 & -36 & -11 & -29 \\ -233 & -119 & 56+i & -35 & 75 & 54 \\ 157 & 81 & -43 & 19+i & -51 & -39 \\ -91 & -48 & 32 & -5 & 30+i & 26 \\ 209 & 107 & -55 & 28 & -69 & -52+i \end{bmatrix}$$

$$\xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 & \frac{1}{5}(7-i) \\ 0 & \boxed{1} & 0 & 0 & 0 & \frac{1}{5}(-9+2i) \\ 0 & 0 & \boxed{1} & 0 & 0 & 1 \\ 0 & 0 & 0 & \boxed{1} & 0 & -1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_F(2-i) = \mathcal{N}(F - (2-i)I_6) = \mathcal{S}p\left(\left\{\begin{bmatrix} \frac{1}{5}(-7+i) \\ \frac{1}{5}(9-2i) \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}\right\}\right)$$

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 complex eigenvalues appearing, 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 [359]. ⊚

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,

$$\lambda = 2 \quad H - 2I_5 = \begin{bmatrix} 13 & 18 & -8 & 6 & -5 \\ 5 & 1 & 1 & -1 & -3 \\ 0 & -4 & 3 & -4 & -2 \\ -43 & -46 & 17 & -16 & 15 \\ 26 & 30 & -12 & 8 & -12 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & -1 \\ 0 & \boxed{1} & 0 & 0 & 1 \\ 0 & 0 & \boxed{1} & 0 & 2 \\ 0 & 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_H(2) = \mathcal{N}(H - 2I_5) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ -1 \\ -2 \\ -1 \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = 1 \quad H - 1I_5 = \begin{bmatrix} 14 & 18 & -8 & 6 & -5 \\ 5 & 2 & 1 & -1 & -3 \\ 0 & -4 & 4 & -4 & -2 \\ -43 & -46 & 17 & -15 & 15 \\ 26 & 30 & -12 & 8 & -11 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & -\frac{1}{2} \\ 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \boxed{1} & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_H(1) = \mathcal{N}(H - 1I_5) = \mathcal{S}p\left(\left\{\begin{bmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \\ -1 \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = 0 \quad H - 0I_5 = \begin{bmatrix} 15 & 18 & -8 & 6 & -5 \\ 5 & 3 & 1 & -1 & -3 \\ 0 & -4 & 5 & -4 & -2 \\ -43 & -46 & 17 & -14 & 15 \\ 26 & 30 & -12 & 8 & -10 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 1 \\ 0 & \boxed{1} & 0 & 0 & -2 \\ 0 & 0 & \boxed{1} & 0 & -2 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_H(0) = \mathcal{N}(H - 0I_5) = \mathcal{S}p\left(\left\{\begin{bmatrix} -1 \\ 2 \\ 2 \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = -1 \quad H + 1I_5 = \begin{bmatrix} 16 & 18 & -8 & 6 & -5 \\ 5 & 4 & 1 & -1 & -3 \\ 0 & -4 & 6 & -4 & -2 \\ -43 & -46 & 17 & -13 & 15 \\ 26 & 30 & -12 & 8 & -9 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & -1/2 \\ 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_H(-1) = \mathcal{N}(H + 1I_5) = \mathcal{S}p\left(\left\{\begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \\ -\frac{1}{2} \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = -3 \quad H + 3I_5 = \begin{bmatrix} 18 & 18 & -8 & 6 & -5 \\ 5 & 6 & 1 & -1 & -3 \\ 0 & -4 & 8 & -4 & -2 \\ -43 & -46 & 17 & -11 & 15 \\ 26 & 30 & -12 & 8 & -7 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & 0 & 0 & -1 \\ 0 & \boxed{1} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & \boxed{1} & 0 & 1 \\ 0 & 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$E_H(-3) = \mathcal{N}(H + 3I_5) = \mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ -\frac{1}{2} \\ -1 \\ -2 \\ 1 \end{bmatrix}\right\}\right)$$

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 [354]. Second, all the eigenvalues are distinct, yielding algebraic and geometric multiplicities of 1 for each eigenvalue, illustrating Theorem DED [375]. ⊚

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 $p(A)$.

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 a square matrix and $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_p\}$ is a set of eigenvectors with eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_p$ 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 [331]), so assume for the remainder that $p \geq 2$.

Suppose to the contrary that S is a linearly dependent set. Define k to be the smallest integer, such that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{k-1}\}$ is linearly independent and $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_k\}$ is linearly dependent. Since eigenvectors are nonzero, the set $\{\mathbf{x}_1\}$ is linearly independent, so $k \geq 2$. Since we are assuming that S is linearly dependent, there is such a k and $k \leq p$. So $2 \leq k \leq p$.

Since $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_k\}$ is linearly dependent there are scalars, $a_1, a_2, a_3, \ldots, a_k$, some non-zero, so that

$$\mathbf{0} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \cdots + a_k\mathbf{x}_k \qquad (*)$$

In particular, we know that $a_k \neq 0$, for if $a_k = 0$, the scalars $a_1, a_2, a_3, \ldots, a_{k-1}$ would include some nonzero values and would give a nontrivial relation of linear dependence on $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{k-1}\}$, contradicting the linear dependence of $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_{k-1}\}$. Now

$$\begin{aligned}
\mathbf{0} &= A\mathbf{0} && \text{Theorem MMZM [200]}\\
&= A\left(a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \cdots + a_k\mathbf{x}_k\right) && \text{Substitute } (*)\\
&= A(a_1\mathbf{x}_1) + A(a_2\mathbf{x}_2) + A(a_3\mathbf{x}_3) + \cdots + A(a_k\mathbf{x}_k) && \text{Theorem MMDAA [201]}\\
&= a_1A\mathbf{x}_1 + a_2A\mathbf{x}_2 + a_3A\mathbf{x}_3 + \cdots + a_kA\mathbf{x}_k && \text{Theorem MMSMM [202]}\\
&= a_1\lambda_1\mathbf{x}_1 + a_2\lambda_2\mathbf{x}_2 + a_3\lambda_3\mathbf{x}_3 + \cdots + a_k\lambda_k\mathbf{x}_k && \mathbf{x}_i \text{ eigenvector of } A \text{ for } \lambda_i \qquad (**)
\end{aligned}$$

Also,

$$\begin{aligned}
\mathbf{0} &= \lambda_k\mathbf{0} && \text{Theorem ZVSM [247]}\\
&= \lambda_k\left(a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 + \cdots + a_k\mathbf{x}_k\right) && \text{Substitute } (*)\\
&= \lambda_k a_1\mathbf{x}_1 + \lambda_k a_2\mathbf{x}_2 + \lambda_k a_3\mathbf{x}_3 + \cdots + \lambda_k a_k\mathbf{x}_k && \text{Distributivity} \qquad (***)
\end{aligned}$$

Put it all together,

$$
\begin{aligned}
\mathbf{0} &= \mathbf{0} - \mathbf{0} && \text{Property of zero vector}\\
&= (a_1\lambda_1\mathbf{x}_1 + a_2\lambda_2\mathbf{x}_2 + a_3\lambda_3\mathbf{x}_3 + \cdots + a_k\lambda_k\mathbf{x}_p)\\
&\quad - (\lambda_k a_1\mathbf{x}_1 + \lambda_k a_2\mathbf{x}_2 + \lambda_k a_3\mathbf{x}_3 + \cdots + \lambda_k a_k\mathbf{x}_k) && \text{Substitute } (**),\ (***)\\
&= (a_1\lambda_1\mathbf{x}_1 - \lambda_k a_1\mathbf{x}_1) + (a_2\lambda_2\mathbf{x}_2 - \lambda_k a_2\mathbf{x}_2) + (a_3\lambda_3\mathbf{x}_3 - \lambda_k a_3\mathbf{x}_3)\\
&\quad + \cdots + (a_{k-1}\lambda_{k-1}\mathbf{x}_{k-1} - \lambda_k a_{k-1}\mathbf{x}_{k-1}) + (a_k\lambda_k\mathbf{x}_k - \lambda_k a_k\mathbf{x}_k) && \text{Commutativity}\\
&= a_1(\lambda_1 - \lambda_k)\mathbf{x}_1 + a_2(\lambda_2 - \lambda_k)\mathbf{x}_2 + a_3(\lambda_3 - \lambda_k)\mathbf{x}_3\\
&\quad + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)\mathbf{x}_{k-1} && \text{Distributivity}
\end{aligned}
$$

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, the eigenvalues were assumed to be distinct, so $\lambda_i \ne \lambda_k$ for $1 \le i \le k-1$. Thus $a_i = 0$ for $1 \le i \le k-1$. Earlier, we deduced that $a_k = 0$ also. Now all these scalars are zero, while they were introduced as having some nonzero values. This is our desired contradiction, and therefore S is linearly independent. ■

There is a simple connection between the eigenvalues of a matrix and whether or not it 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:

$$
\begin{aligned}
A \text{ is singular} &\iff \text{there exists } \mathbf{x} \ne \mathbf{0},\ A\mathbf{x} = \mathbf{0} && \text{Definition NSM [73]}\\
&\iff \text{there exists } \mathbf{x} \ne \mathbf{0},\ A\mathbf{x} = 0\mathbf{x} && \text{Theorem MMZM [200]}\\
&\iff \lambda = 0 \text{ is an eigenvalue of } A && \text{Definition EEM [331]}
\end{aligned}
$$
■

With an equivalence about singular matrices we can update our list of equivalences about nonsingular matrices.

Theorem NSME8
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. The range of A is $\mathbb{C}^n$, $\mathcal{R}(A) = \mathbb{C}^n$.
7. A is invertible.
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 [354]. ■

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

$$\begin{aligned} (\alpha A)\mathbf{x} &= \alpha(A\mathbf{x}) && \text{Theorem MMSMM [202]} \\ &= \alpha(\lambda\mathbf{x}) && \mathbf{x} \text{ eigenvector of } A \\ &= (\alpha\lambda)\mathbf{x} && \text{Associativity} \end{aligned}$$

So $\mathbf{x} \neq 0$ 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.

Theorem EOMP
Eigenvalues Of Matrix Powers
Suppose A is a square matrix, λ is an eigenvalue of A, and $s \geq 0$ 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. First, for $s = 0$,

$$\begin{aligned} A^s\mathbf{x} &= A^0\mathbf{x} \\ &= I_n\mathbf{x} \\ &= \mathbf{x} && \text{Theorem MMIM [201]} \\ &= 1\mathbf{x} && \text{Property of 1} \\ &= \lambda^0\mathbf{x} \\ &= \lambda^s\mathbf{x} \end{aligned}$$

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

$$\begin{aligned}
A^{s+1}\mathbf{x} &= A^s A\mathbf{x} \\
&= A^s(\lambda\mathbf{x}) && \mathbf{x} \text{ eigenvector of } A \text{ for } \lambda \\
&= \lambda(A^s\mathbf{x}) && \text{Theorem MMSMM [202]} \\
&= \lambda(\lambda^s\mathbf{x}) && \text{Induction hypothesis} \\
&= (\lambda\lambda^s)\mathbf{x} && \text{Associativity} \\
&= \lambda^{s+1}\mathbf{x}
\end{aligned}$$

So $\mathbf{x} \neq \mathbf{0}$ is an eigenvector of A^{s+1} for λ^{s+1}, and induction (Technique XX [**??**]) tells us the theorem is true for all $s \geq 0$. ■

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 [335] and the characteristic polynomial (Definition CP [339]). 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 $q(x) = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$. Then

$$\begin{aligned}
q(A)\mathbf{x} &= \left(a_0A^0 + a_1A^1 + a_2A^2 + \cdots + a_mA^m\right)\mathbf{x} \\
&= (a_0A^0)\mathbf{x} + (a_1A^1)\mathbf{x} + (a_2A^2)\mathbf{x} + \cdots + (a_mA^m)\mathbf{x} && \text{Theorem MMDAA [201]} \\
&= a_0(A^0\mathbf{x}) + a_1(A^1\mathbf{x}) + a_2(A^2\mathbf{x}) + \cdots + a_m(A^m\mathbf{x}) && \text{Theorem MMSMM [202]} \\
&= a_0(\lambda^0\mathbf{x}) + a_1(\lambda^1\mathbf{x}) + a_2(\lambda^2\mathbf{x}) + \cdots + a_m(\lambda^m\mathbf{x}) && \text{Theorem EOMP [355]} \\
&= (a_0\lambda^0)\mathbf{x} + (a_1\lambda^1)\mathbf{x} + (a_2\lambda^2)\mathbf{x} + \cdots + (a_m\lambda^m)\mathbf{x} && \text{Associativity} \\
&= \left(a_0\lambda^0 + a_1\lambda^1 + a_2\lambda^2 + \cdots + a_m\lambda^m\right)\mathbf{x} && \text{Distributivity} \\
&= q(\lambda)\mathbf{x}
\end{aligned}$$

So $\mathbf{x} \neq 0$ is an eigenvector of $q(A)$ for the eigenvalue $q(\lambda)$. ■

Example BDE
Building desired eigenvalues
In Example ESMS4 [345] 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 [356] 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 discus how to concoct this polynomial, but a text on numerical analysis should provide the details. In our case, 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 [356] 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 NSI [228], but more importantly, $\frac{1}{\lambda}$ does not involve division by zero since Theorem SMZE [354] prohibits this possibility.

Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for λ. Suppose A has size n. Then

$$\begin{aligned}
A^{-1}\mathbf{x} &= A^{-1}(1\mathbf{x}) && \text{Property of 1}\\
&= A^{-1}(\frac{1}{\lambda}\lambda\mathbf{x}) &&\\
&= \frac{1}{\lambda}A^{-1}(\lambda\mathbf{x}) && \text{Theorem MMSMM [202]}\\
&= \frac{1}{\lambda}A^{-1}(A\mathbf{x}) && \mathbf{x}\text{ eigenvector of }A\\
&= \frac{1}{\lambda}(A^{-1}A)\mathbf{x} && \text{Associativity}\\
&= \frac{1}{\lambda}I_n\mathbf{x} && \text{Definition MI [216]}\\
&= \frac{1}{\lambda}\mathbf{x} && \text{Theorem MMIM [201]}
\end{aligned}$$

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 Let $\mathbf{x} \neq \mathbf{0}$ be one eigenvector of A for λ. Suppose A has size n. Then

$$\begin{aligned}
p_A(x) &= \det(A - xI_n) && \text{Definition CP [339]}\\
&= \det\left((A - xI_n)^t\right) && \text{Theorem DT [327]}\\
&= \det\left(A^t - (xI_n)^t\right) && \text{Theorem TASM [164]}\\
&= \det\left(A^t - xI_n^t\right) && \text{Theorem TASM [164]}\\
&= \det\left(A^t - xI_n\right) && \text{Definition IM [76]}\\
&= p_{A^t}(x) && \text{Definition CP [339]}
\end{aligned}$$

So A and A^t have the same characteristic polynomial, and by Theorem EMRCP [340], their eigenvalues are identical and have equal algebraic multiplicities. Notice that what we have proved here is a bit stronger than the conclusion of the theorem. ■

If a matrix has only real entries, then the computation of the characteristic polynomial (Definition CP [339]) 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

$$\begin{aligned} A\overline{\mathbf{x}} &= \overline{A}\overline{\mathbf{x}} && A \text{ has real entries} \\ &= \overline{A\mathbf{x}} && \text{Theorem MMCC [204]} \\ &= \overline{\lambda\mathbf{x}} && \mathbf{x} \text{ eigenvector of } A \\ &= \overline{\lambda}\overline{\mathbf{x}} && \text{Theorem CCRSM [146]} \end{aligned}$$

So $\overline{\mathbf{x}}$ is an eigenvector of A for the eigenvalue $\overline{\lambda}$. ■

This phenomenon is amply illustrated in Example CEMS6 [346], 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 [359] 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 TODO: ■

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 [343]), we can factor the characteristic polynomial as

$$p_A(x) = (x-\lambda_1)^{\alpha_A(\lambda_1)}(x-\lambda_2)^{\alpha_A(\lambda_2)}(x-\lambda_3)^{\alpha_A(\lambda_3)}\cdots(x-\lambda_k)^{\alpha_A(\lambda_k)}$$

The left-hand side is a polynomial of degree n by Theorem DCP [359] and the right-hand side is a polynomial of degree $\sum_{i=1}^{k}\alpha_A(\lambda_i) = n$, so the equality of the polynomials' degrees gives the result. ■

Theorem ME
Multiplicities of an Eigenvalue
Suppose that A is a square matrix of size n and λ is an eigenvalue. Then

$$1 \leq \gamma_A(\lambda) \leq \alpha_A(\lambda) \leq n$$ □

Proof Since λ is an eigenvalue of A, there is an eigenvector of A for λ, $\mathbf{x}$. Then $\mathbf{x} \in E_A(\lambda)$, so $\gamma_A(\lambda) \geq 1$, since we can extend $\{\mathbf{x}\}$ into a basis of $E_A(\lambda)$.

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, $E_A(\lambda)$. 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 [309]. Finally, define a matrix S by

$$S = [\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\ldots|\mathbf{x}_g|\mathbf{y}_1|\mathbf{y}_2|\mathbf{y}_3|\ldots|\mathbf{y}_{n-g}] = [\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\ldots|\mathbf{x}_g|R]$$

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 NSLIC [137]) and therefore invertible (Theorem NSI [228]). Then,

$$\begin{aligned}
[\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\ldots|\mathbf{e}_n] &= I_n\\
&= S^{-1}S\\
&= S^{-1}[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\ldots|\mathbf{x}_g|R]\\
&= [S^{-1}\mathbf{x}_1|S^{-1}\mathbf{x}_2|S^{-1}\mathbf{x}_3|\ldots|S^{-1}\mathbf{x}_g|S^{-1}R]
\end{aligned}$$

So

$$S^{-1}\mathbf{x}_i = \mathbf{e}_i \quad 1 \leq i \leq g \qquad (*)$$

Preparations in place, we compute the characteristic polynomial of A,

$$\begin{aligned}
p_A(x) &= \det(A - xI_n) && \text{Definition CP [339]}\\
&= 1\det(A - xI_n) \\
&= \det(I_n)\det(A - xI_n) && \text{Definition DM [323]}\\
&= \det\left(S^{-1}S\right)\det(A - xI_n) && \text{Definition MI [216]}\\
&= \det\left(S^{-1}\right)\det(S)\det(A - xI_n) && \text{Theorem DRMM [328]}\\
&= \det\left(S^{-1}\right)\det(A - xI_n)\det(S) && \text{Commutativity in } \mathbb{C}\\
&= \det\left(S^{-1}(A - xI_n)S\right) && \text{Theorem DRMM [328]}\\
&= \det\left(S^{-1}AS - S^{-1}xI_nS\right) && \text{Theorem MMDAA [201]}\\
&= \det\left(S^{-1}AS - xS^{-1}I_nS\right) && \text{Theorem MMSMM [202]}\\
&= \det\left(S^{-1}AS - xS^{-1}S\right) && \text{Theorem MMIM [201]}\\
&= \det\left(S^{-1}AS - xI_n\right) && \text{Definition MI [216]}\\
&= p_{S^{-1}AS}(x) && \text{Definition CP [339]}
\end{aligned}$$

What can we learn then about the matrix $S^{-1}AS$?

$$\begin{aligned}
S^{-1}AS &= S^{-1}A[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\dots|\mathbf{x}_g|R]\\
&= S^{-1}[A\mathbf{x}_1|A\mathbf{x}_2|A\mathbf{x}_3|\dots|A\mathbf{x}_g|AR] && \text{Definition MM [197]}\\
&= S^{-1}[\lambda\mathbf{x}_1|\lambda\mathbf{x}_2|\lambda\mathbf{x}_3|\dots|\lambda\mathbf{x}_g|AR] && \mathbf{x}_i \text{ eigenvectors of } A\\
&= [S^{-1}\lambda\mathbf{x}_1|S^{-1}\lambda\mathbf{x}_2|S^{-1}\lambda\mathbf{x}_3|\dots|S^{-1}\lambda\mathbf{x}_g|S^{-1}AR] && \text{Definition MM [197]}\\
&= [\lambda S^{-1}\mathbf{x}_1|\lambda S^{-1}\mathbf{x}_2|\lambda S^{-1}\mathbf{x}_3|\dots|\lambda S^{-1}\mathbf{x}_g|S^{-1}AR] && \text{Theorem MMSMM [202]}\\
&= [\lambda\mathbf{e}_1|\lambda\mathbf{e}_2|\lambda\mathbf{e}_3|\dots|\lambda\mathbf{e}_g|S^{-1}AR] && S^{-1}S = I_n, \text{ ((}* \text{ above)}
\end{aligned}$$

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 $(x - \lambda)$. More precisely, let T be the square matrix of size $n - g$ that is formed from the last $n - g$ rows of $S^{-1}AR$. Then

$$p_A(x) = p_{S^{-1}AS}(x) = (x - \lambda)^g p_T(x).$$

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 [343]). In other words,

$$\gamma_A(\lambda) = g \leq \alpha_A(\lambda)$$

as desired.

Theorem NEM [359] 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. □

Proof Suppose that A has k distinct eigenvalues, $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$. Then

$$\begin{aligned} k &= \sum_{i=1}^{k} 1 \\ &\leq \sum_{i=1}^{k} \alpha_A(\lambda) && \text{Theorem ME [360]} \\ &= n && \text{Theorem NEM [359]} \end{aligned}$$ ■

Subsection EHM
Eigenvalues of Hermitian Matrices

Recall that a matrix is Hermitian (or self-adjoint) if $A = \left(\overline{A}\right)^t$ (Definition HM [232]). In the case where A is a matrix whose entries are all real numbers, being Hermitian is identical to being symmetric (Definition SYM [163]). 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 λ. Then

$$\begin{aligned} \lambda \langle \mathbf{x},\, \mathbf{x} \rangle &= \langle \lambda\mathbf{x},\, \mathbf{x} \rangle && \text{Theorem IPSM [148]} \\ &= \langle A\mathbf{x},\, \mathbf{x} \rangle && \mathbf{x} \text{ eigenvector of } A \\ &= (A\mathbf{x})^t\, \overline{\mathbf{x}} && \text{Theorem MMIP [203]} \\ &= \mathbf{x}^t A^t \overline{\mathbf{x}} && \text{Theorem MMT [205]} \\ &= \mathbf{x}^t \left(\left(\overline{A}\right)^t\right)^t \overline{\mathbf{x}} && \text{Definition HM [232]} \\ &= \mathbf{x}^t \overline{A}\overline{\mathbf{x}} && \text{Theorem TASM [164]} \\ &= \mathbf{x}^t \overline{A\mathbf{x}} && \text{Theorem MMCC [204]} \\ &= \langle \mathbf{x},\, A\mathbf{x} \rangle && \text{Theorem MMIP [203]} \\ &= \langle \mathbf{x},\, \lambda\mathbf{x} \rangle && \mathbf{x} \text{ eigenvector of } A \\ &= \overline{\lambda} \langle \mathbf{x},\, \mathbf{x} \rangle && \text{Theorem IPSM [148]} \end{aligned}$$

Since $\mathbf{x} \neq \mathbf{0}$, Theorem PIP [150] says that $\langle \mathbf{x}, \mathbf{x} \rangle \neq 0$, so we can "cancel" $\langle \mathbf{x}, \mathbf{x} \rangle$ from both sides of this equality. This leaves $\lambda = \overline{\lambda}$, so λ has a complex part equal to zero, and therefore is a real number. ■

Look back and compare Example ESMS4 [345] and Example CEMS6 [346]. In Example CEMS6 [346] 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 [345], the matrix has only real entries, but is also symmetric. So by Theorem HMRE [362], 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 [362] 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 $\mathbf{x} \neq \mathbf{0}$ be an eigenvector of A for λ and let $\mathbf{y} \neq \mathbf{0}$ be an eigenvector of A for ρ. By Theorem HMRE [362], we know that ρ must be a real number. Then

$$\begin{aligned}
\lambda \langle \mathbf{x}, \mathbf{y} \rangle &= \langle \lambda\mathbf{x}, \mathbf{y} \rangle && \text{Theorem IPSM [148]}\\
&= \langle A\mathbf{x}, \mathbf{y} \rangle && \mathbf{x} \text{ eigenvector of } A\\
&= (A\mathbf{x})^t\, \overline{\mathbf{y}} && \text{Theorem MMIP [203]}\\
&= \mathbf{x}^t A^t \overline{\mathbf{y}} && \text{Theorem MMT [205]}\\
&= \mathbf{x}^t \left(\left(\overline{A}\right)^t\right)^t \overline{\mathbf{y}} && \text{Definition HM [232]}\\
&= \mathbf{x}^t \overline{A}\,\overline{\mathbf{y}} && \text{Theorem TASM [164]}\\
&= \mathbf{x}^t \overline{A\mathbf{y}} && \text{Theorem XX [??], conj. matrix mult}\\
&= \langle \mathbf{x}, A\mathbf{y} \rangle && \text{Theorem MMIP [203]}\\
&= \langle \mathbf{x}, \rho\mathbf{y} \rangle && \mathbf{y} \text{ eigenvector of } A\\
&= \overline{\rho} \langle \mathbf{x}, \mathbf{y} \rangle && \text{Theorem IPSM [148]}\\
&= \rho \langle \mathbf{x}, \mathbf{y} \rangle && \text{Theorem HMRE [362]}
\end{aligned}$$

Since $\lambda \neq \rho$, we conclude that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ and so $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors (Definition OV [151]). ■

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?

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 [441].

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 [441] 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 [367] 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, lets 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

$$\begin{aligned}
A &= S^{-1}BS \\
&= \begin{bmatrix} 10 & 1 & 0 & 2 & -5 \\ -1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 2 & 1 & -3 \\ 0 & 0 & -1 & 0 & 1 \\ -4 & -1 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} -4 & 1 & -3 & -2 & 2 \\ 1 & 2 & -1 & 3 & -2 \\ -4 & 1 & 3 & 2 & 2 \\ -3 & 4 & -2 & -1 & -3 \\ 3 & 1 & -1 & 1 & -4 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 1 & 1 \\ 0 & 1 & -1 & -2 & -1 \\ 1 & 3 & -1 & 1 & 1 \\ -2 & -3 & 3 & 1 & -2 \\ 1 & 3 & -1 & 2 & 1 \end{bmatrix} \\
&= \begin{bmatrix} -10 & -27 & -29 & -80 & -25 \\ -2 & 6 & 6 & 10 & -2 \\ -3 & 11 & -9 & -14 & -9 \\ -1 & -13 & 0 & -10 & -1 \\ 11 & 35 & 6 & 49 & 19 \end{bmatrix}
\end{aligned}$$

So by this construction, we know that A and B are similar. ⊚

Let's do that again.

Example SMS4
Similar matrices of size 4
Define

$$B = \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \qquad S = \begin{bmatrix} 1 & 1 & 2 \\ -2 & -1 & -3 \\ 1 & -2 & 0 \end{bmatrix}$$

Check that S is nonsingular and then compute

$$\begin{aligned}
A &= S^{-1}BS \\
&= \begin{bmatrix} -6 & -4 & -1 \\ -3 & -2 & -1 \\ 5 & 3 & 1 \end{bmatrix} \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ -2 & -1 & -3 \\ 1 & -2 & 0 \end{bmatrix} \\
&= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}
\end{aligned}$$

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 [327], $\det(A) = (-1)(3)(-1) = 3$. 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 [218]). ⊚

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 [48]). 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 NSRRI [76]), 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 [365]. By Theorem MIMI [224], S^{-1} is invertible, and by Theorem NSI [228] is therefore nonsingular. So

$$\begin{aligned}
(S^{-1})^{-1}A(S^{-1}) &= SAS^{-1} && \text{Theorem MIMI [224]}\\
&= SS^{-1}BSS^{-1} && \text{Substitution for } A\\
&= \left(SS^{-1}\right)B\left(SS^{-1}\right) && \text{Theorem MMA [202]}\\
&= I_nBI_n && \text{Definition MI [216]}\\
&= B && \text{Theorem MMIM [201]}
\end{aligned}$$

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 [365]. (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 [223] and then nonsingular by Theorem NSI [228]. Now

$$\begin{aligned}
(RS)^{-1}C(RS) &= S^{-1}R^{-1}CRS && \text{Theorem SS [223]}\\
&= S^{-1}\left(R^{-1}CR\right)S && \text{Theorem MMA [202]}\\
&= S^{-1}BS && \text{Substitution of } B\\
&= A
\end{aligned}$$

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 Suppose A and B have size n and are similar via the nonsingular matrix S, so $A = S^{-1}BS$ by Definition SIM [365].

$$\begin{aligned}
p_A(x) &= \det(A - xI_n) && \text{Definition CP [339]}\\
&= \det\left(S^{-1}BS - xI_n\right) && \text{Substitution for } A\\
&= \det\left(S^{-1}BS - xS^{-1}I_nS\right) && \text{Theorem MMIM [201]}\\
&= \det\left(S^{-1}BS - S^{-1}xI_nS\right) && \text{Theorem MMSMM [202]}\\
&= \det\left(S^{-1}(B - xI_n)S\right) && \text{Theorem MMDAA [201]}\\
&= \det\left(S^{-1}\right)\det(B - xI_n)\det(S) && \text{Theorem DRMM [328]}\\
&= \det\left(S^{-1}\right)\det(S)\det(B - xI_n) && \text{Commutativity in } \mathbb{C}\\
&= \det\left(S^{-1}S\right)\det(B - xI_n) && \text{Theorem DRMM [328]}\\
&= \det(I_n)\det(B - xI_n) && \text{Definition MI [216]}\\
&= 1\det(B - xI_n) && \text{Definition DM [323]}\\
&= p_B(x) && \text{Definition CP [339]}
\end{aligned}$$

■

So similar matrices not only have the same *set* of eigenvalues, the algebraic multiplicities 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

$$A = \begin{bmatrix}1 & 1\\0 & 1\end{bmatrix} \qquad B = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

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 [368] was true, then A and B would be similar. Suppose this was the case. In other words, there is a nonsingular matrix S so that $A = S^{-1}BS$. Then

$$A = S^{-1}BS = S^{-1}I_2S = S^{-1}S = I_2 \neq A$$

this contradiction tells us that the converse of Theorem SMEE [368] 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 $A = (a_{ij})$ 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. △

Example DAB
Diagonalization of Archetype B
Archetype B [478] 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,

$$\begin{aligned} S^{-1}BS &= \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & -1 & -1 \\ 2 & 3 & 1 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} -7 & -6 & -12 \\ 5 & 5 & 7 \\ 1 & 0 & 4 \end{bmatrix} \begin{bmatrix} -5 & -3 & -2 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \end{aligned}$$

⊚

Example SMS4 [366] 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 [478] and compute the eigenvalues and eigenvectors of the matrix in Example SMS4 [366].

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. □

Proof (⇒) Let $S = \{\mathbf{x}_1,\,\mathbf{x}_2,\,\mathbf{x}_3,\,\ldots,\,\mathbf{x}_n\}$ be a linearly independent set of eigenvectors of A for the eigenvalues $\lambda_1,\,\lambda_2,\,\lambda_3,\,\ldots,\,\lambda_n$. Recall Definition SUV [218] and define

$$\begin{aligned} R &= [\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\ldots|\mathbf{x}_n] \\ D &= \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\lambda_1\mathbf{e}_1|\lambda_2\mathbf{e}_2|\lambda_3\mathbf{e}_3|\ldots|\lambda_n\mathbf{e}_n] \end{aligned}$$

The columns of R are the vectors of the linearly independent set S and so by Theorem NSLIC [137] the matrix R is nonsingular. By Theorem NSI [228] we know R^{-1}

exists.

$$\begin{aligned}
R^{-1}AR &= R^{-1}A\left[\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3|\dots|\mathbf{x}_n\right] \\
&= R^{-1}[A\mathbf{x}_1|A\mathbf{x}_2|A\mathbf{x}_3|\dots|A\mathbf{x}_n] && \text{Definition MM [197]} \\
&= R^{-1}[\lambda_1\mathbf{x}_1|\lambda_2\mathbf{x}_2|\lambda_3\mathbf{x}_3|\dots|\lambda_n\mathbf{x}_n] && \mathbf{x}_i \text{ eigenvector of } A \text{ for } \lambda_i \\
&= R^{-1}[\lambda_1 R\mathbf{e}_1|\lambda_2 R\mathbf{e}_2|\lambda_3 R\mathbf{e}_3|\dots|\lambda_n R\mathbf{e}_n] && \text{Definition MVP [195]} \\
&= R^{-1}[R(\lambda_1\mathbf{e}_1)|R(\lambda_2\mathbf{e}_2)|R(\lambda_3\mathbf{e}_3)|\dots|R(\lambda_n\mathbf{e}_n)] && \text{Theorem MMSMM [202]} \\
&= R^{-1}R[\lambda_1\mathbf{e}_1|\lambda_2\mathbf{e}_2|\lambda_3\mathbf{e}_3|\dots|\lambda_n\mathbf{e}_n] && \text{Definition MM [197]} \\
&= I_nD && \text{Definition MI [216]} \\
&= D && \text{Theorem MMIM [201]}
\end{aligned}$$

This says that A is similar to the diagonal matrix D via the nonsingular matrix R. Thus A is diagonalizable (Definition DZM [369]).

($\Leftarrow$) 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 [218])

$$E = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix} = [d_1\mathbf{e}_1|d_2\mathbf{e}_2|d_3\mathbf{e}_3|\dots|d_n\mathbf{e}_n]$$

such that $T^{-1}AT = E$.

$$\begin{aligned}
[A\mathbf{y}_1|A\mathbf{y}_2|A\mathbf{y}_3|\dots|A\mathbf{y}_n] &= A\left[\mathbf{y}_1|\mathbf{y}_2|\mathbf{y}_3|\dots|\mathbf{y}_n\right] && \text{Definition MM [197]} \\
&= AT \\
&= I_nAT && \text{Theorem MMIM [201]} \\
&= TT^{-1}AT && \text{Definition MI [216]} \\
&= TE && \text{Substitution} \\
&= T[d_1\mathbf{e}_1|d_2\mathbf{e}_2|d_3\mathbf{e}_3|\dots|d_n\mathbf{e}_n] \\
&= [T(d_1\mathbf{e}_1)|T(d_2\mathbf{e}_2)|T(d_3\mathbf{e}_3)|\dots|T(d_n\mathbf{e}_n)] && \text{Definition MM [197]} \\
&= [d_1T\mathbf{e}_1|d_2T\mathbf{e}_2|d_3T\mathbf{e}_3|\dots|d_nT\mathbf{e}_n] && \text{Definition MM [197]} \\
&= [d_1\mathbf{y}_1|d_2\mathbf{y}_2|d_3\mathbf{y}_3|\dots|d_n\mathbf{y}_n] && \text{Definition MVP [195]}
\end{aligned}$$

This equality of matrices allows us to conclude that the columns are equal vectors. That is, $A\mathbf{y}_i = d_i\mathbf{y}_i$ for $1 \leq i \leq n$. In other words, $\mathbf{y}_i$ is an eigenvector of A for the eigenvalue d_i. (Why can't $\mathbf{y}_i = \mathbf{0}$?). Because T is nonsingular, the set containing T's columns, $S = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \dots, \mathbf{y}_n\}$, is a linearly independent set (Theorem NSLIC [137]). So the set S has all the required properties. ∎

Notice that the proof of Theorem DC [370] 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 [340], Example EMS3 [340] and Example ESMS3 [342]. F's eigenvalues and eigenspaces are

$$\lambda = 3 \qquad E_F(3) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix}\right\}\right)$$

$$\lambda = -1 \qquad E_F(-1) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{2}{3} \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{3} \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

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 NSRRI [76] or has a nonzero determinant, Theorem SMZD [328]). Then the three columns of S are a linearly independent set (Theorem NSLIC [137]). By Theorem DC [370] we now know that F is diagonalizable. Furthermore, the construction in the proof of Theorem DC [370] 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,

$$\begin{aligned} S^{-1}FS &= \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 6 & 4 & 2 \\ -3 & -1 & -1 \\ -6 & -4 & -1 \end{bmatrix} \begin{bmatrix} -13 & -8 & -4 \\ 12 & 7 & 4 \\ 24 & 16 & 7 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & -\frac{2}{3} & -\frac{1}{3} \\ \frac{1}{2} & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \end{aligned}$$

Note that the above computations can be viewed two ways. The proof of Theorem DC [370] tells us that the four matrices (F, S, F^{-1} 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 [360]. When every eigenvalue's eigenspace is this big, then we can diagonalize the matrix, and only then. Three examples we have seen so far in this section, Example SMS5 [365], Example DAB [369] and Example DMS3 [371], 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 DMLE
Diagonalizable Matrices have Large 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, $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$.

($\Leftarrow$) Let $S_i = \left\{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3}, \ldots, \mathbf{x}_{i\gamma_A(\lambda_i)}\right\}$, be a basis for the eigenspace of λ_i, $E_A(\lambda_i)$, $1 \leq i \leq k$. 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 (why not?) so the sets S_i have no vectors in common. Thus the size of S is

$$\begin{aligned} \sum_{i=1}^{k} \gamma_A(\lambda_i) &= \sum_{i=1}^{k} \alpha_A(\lambda_i) && \text{Hypothesis} \\ &= n && \text{Theorem NEM [359]} \end{aligned}$$

We now want to show that S is a linearly independent set. So we will begin with a relation of linear dependence on S, using doubly-subscripted eigenvectors,

$$\begin{aligned} \mathbf{0} = &\left(a_{11}\mathbf{x}_{11} + a_{12}\mathbf{x}_{12} + \cdots + a_{1\gamma_A(\lambda_1)}\mathbf{x}_{1\gamma_A(\lambda_1)}\right) + \left(a_{21}\mathbf{x}_{21} + a_{22}\mathbf{x}_{22} + \cdots + a_{2\gamma_A(\lambda_2)}\mathbf{x}_{2\gamma_A(\lambda_2)}\right) \\ &+ \cdots + \left(a_{k1}\mathbf{x}_{k1} + a_{k2}\mathbf{x}_{k2} + \cdots + a_{k\gamma_A(\lambda_k)}\mathbf{x}_{k\gamma_A(\lambda_k)}\right) \end{aligned}$$

Define the vectors $\mathbf{y}_i$, $1 \leq i \leq k$ by

$$\begin{aligned} \mathbf{y}_1 &= \left(a_{11}\mathbf{x}_{11} + a_{12}\mathbf{x}_{12} + a_{13}\mathbf{x}_{13} + \cdots + a_{\gamma_A(1\lambda_1)}\mathbf{x}_{1\gamma_A(\lambda_1)}\right) \\ \mathbf{y}_2 &= \left(a_{21}\mathbf{x}_{21} + a_{22}\mathbf{x}_{22} + a_{23}\mathbf{x}_{23} + \cdots + a_{\gamma_A(2\lambda_2)}\mathbf{x}_{2\gamma_A(\lambda_2)}\right) \\ \mathbf{y}_3 &= \left(a_{31}\mathbf{x}_{31} + a_{32}\mathbf{x}_{32} + a_{33}\mathbf{x}_{33} + \cdots + a_{\gamma_A(3\lambda_3)}\mathbf{x}_{3\gamma_A(\lambda_3)}\right) \\ &\vdots \\ \mathbf{y}_k &= \left(a_{k1}\mathbf{x}_{k1} + a_{k2}\mathbf{x}_{k2} + a_{k3}\mathbf{x}_{k3} + \cdots + a_{\gamma_A(k\lambda_k)}\mathbf{x}_{k\gamma_A(\lambda_k)}\right) \end{aligned}$$

Then the relation of linear dependence becomes

$$\mathbf{0} = \mathbf{y}_1 + \mathbf{y}_2 + \mathbf{y}_3 + \cdots + \mathbf{y}_k$$

Since the eigenspace $E_A(\lambda_i)$ is closed under vector addition and scalar multiplication, $\mathbf{y}_i \in E_A(\lambda_i)$, $1 \leq i \leq k$. 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 [353]. 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 [353]. 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 $E_A(\lambda_i)$, which is therefore linearly independent. From these relations of linear dependence on linearly independent sets we conclude that $a_{ij} = 0$, $1 \leq j \leq \gamma_A(\lambda_i)$ for $1 \leq i \leq k$. This establishes that our original relation of linear dependence on S has only the trivial solution, 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 [370] is diagonalizable.

($\Rightarrow$) Now we assume that A is diagonalizable. Aiming for a contradiction, 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 [360] 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 [370] 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. S is a linearly independent set, so the the subset S_i containing the n_i eigenvectors for λ_i must also be linearly independent. Because the eigenspace $E_A(\lambda_i)$ has dimension $\gamma_A(\lambda_i)$ and S_i is a linearly independent subset in $E_A(\lambda_i)$, $n_i \leq \gamma_A(\lambda_i)$, $1 \leq i \leq k$. Now,

$$\begin{aligned}
n &= n_1 + n_2 + n_3 + \cdots + n_t + \cdots + n_k && \text{Size of } S\\
&\leq \gamma_A(\lambda_1) + \gamma_A(\lambda_2) + \gamma_A(\lambda_3) + \cdots + \gamma_A(\lambda_t) + \cdots + \gamma_A(\lambda_k) && S_i \text{ linearly independent}\\
&< \alpha_A(\lambda_1) + \alpha_A(\lambda_2) + \alpha_A(\lambda_3) + \cdots + \alpha_A(\lambda_t) + \cdots + \alpha_A(\lambda_k) && \text{Assumption about } \lambda_t\\
&= n && \text{Theorem NEM [359]}
\end{aligned}$$

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 [332], Example CAEHW [337], Example ESMS3 [342], Example ESMS4 [345], Example DEMS5 [349], Archetype B [478], Archetype F [495], Archetype K [520] and Archetype L [525] are all examples of matrices that are diagonalizable and that illustrate Theorem DMLE [373]. While we have provide 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 [344] 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

$$E_B(2) = \mathcal{S}p\left(\left\{\begin{bmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \\ 1 \end{bmatrix}\right\}\right)$$

So the geometric multiplicity of $\lambda = 2$ is $\gamma_B(2) = 1$, while the algebraic multiplicity is $\alpha_B(2) = 3$. By Theorem DMLE [373], the matrix B is not diagonalizable. ⊚

Archetype A [473] 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 [346] 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 [346] which has two complex eigenvalues, each with differing multiplicities. Likewise, Example EMMS4 [344] 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 [359] we have $n = \sum_{i=1}^{n} \alpha_A(\lambda_i)$, which implies that $\alpha_A(\lambda_i) = 1$, $1 \leq i \leq n$. From Theorem ME [360] it follows that $\gamma_A(\lambda_i) = 1$, $1 \leq i \leq n$. So $\gamma_A(\lambda_i) = \alpha_A(\lambda_i)$, $1 \leq i \leq n$ and Theorem DMLE [373] says A is diagonalizable. ■

Example DEHD
Distinct eigenvalues, hence diagonalizable
In Example DEMS5 [349] 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 [375] 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 some of the eigenvectors given in Example DEMS5 [349] as basis vectors for the eigenspaces have been multiplied by suitable nonzero scalars to clear out fractions. These are of course still eigenvectors for the respective eigenvalues by Theorem EMS [341].

$$
\begin{aligned}
&S^{-1}HS\\
&= \begin{bmatrix} 2 & 1 & -1 & 1 & 1 \\ -1 & 0 & 2 & 0 & -1 \\ -2 & 0 & 2 & -1 & -2 \\ -4 & -1 & 0 & -2 & -1 \\ 2 & 2 & 1 & 2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 15 & 18 & -8 & 6 & -5 \\ 5 & 3 & 1 & -1 & -3 \\ 0 & -4 & 5 & -4 & -2 \\ -43 & -46 & 17 & -14 & 15 \\ 26 & 30 & -12 & 8 & -10 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 & 1 & 1 \\ -1 & 0 & 2 & 0 & -1 \\ -2 & 0 & 2 & -1 & -2 \\ -4 & -1 & 0 & -2 & -1 \\ 2 & 2 & 1 & 2 & 1 \end{bmatrix}\\
&= \begin{bmatrix} -3 & -3 & 1 & -1 & 1 \\ -1 & -2 & 1 & 0 & 1 \\ -5 & -4 & 1 & -1 & 2 \\ 10 & 10 & -3 & 2 & -4 \\ -7 & -6 & 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 15 & 18 & -8 & 6 & -5 \\ 5 & 3 & 1 & -1 & -3 \\ 0 & -4 & 5 & -4 & -2 \\ -43 & -46 & 17 & -14 & 15 \\ 26 & 30 & -12 & 8 & -10 \end{bmatrix} \begin{bmatrix} 2 & 1 & -1 & 1 & 1 \\ -1 & 0 & 2 & 0 & -1 \\ -2 & 0 & 2 & -1 & -2 \\ -4 & -1 & 0 & -2 & -1 \\ 2 & 2 & 1 & 2 & 1 \end{bmatrix}\\
&= \begin{bmatrix} -3 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}
\end{aligned}
$$

⊚

Archetype B [478] is another example of a matrix that has as many distinct eigenvalues as its size, and is hence diagonalizable by Theorem DED [375].

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.

TODO: a computational example of this

Subsection OD Orthonormal Diagonalization

Theorem ODHM
Orthonormal Diagonalization of Hermitian Matrices
Suppose that A is a Hermitian matrix of size n. Then A can be diagonalized by a

similarity transformation using an orthonormal transformation. □

Proof TODO: ■

Subsection READ
Reading Questions

1. What is an equivalence relation?
2. When is a matrix diagonalizable?
3. Find a diagonal matrix similar to

$$A = \begin{bmatrix} -5 & 8 \\ -4 & 7 \end{bmatrix}$$

LT: Linear Transformations

Section 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 [239]), the ten axioms, basic properties 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.

Subsection LT
Linear Transformations

Here's a key definition.

Definition LT
Linear Transformation
A **linear transformation**, $T\colon U \mapsto V$, 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 $\alpha \in \mathbb{C}$ △

The two defining conditions of the definition of a linear transformations should "feel linear," whatever that means. Conversely, these two conditions could be taken as a *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.

$$\begin{array}{ccc} \mathbf{u}_1,\,\mathbf{u}_2 & \xrightarrow{T} & T(\mathbf{u}_1)\,,\,T(\mathbf{u}_2) \\ +\big\downarrow & & \big\downarrow + \\ \mathbf{u}_1+\mathbf{u}_2 & \xrightarrow{T} & \begin{array}{c} T(\mathbf{u}_1)+T(\mathbf{u}_2), \\ T(\mathbf{u}_1+\mathbf{u}_2) \end{array} \end{array}$$

$$\begin{array}{ccc} \mathbf{u} & \xrightarrow{T} & T(\mathbf{u}) \\ \alpha\big\downarrow & & \big\downarrow \alpha \\ \alpha\mathbf{u} & \xrightarrow{T} & \begin{array}{c} \alpha T(\mathbf{u}), \\ T(\alpha\mathbf{u}) \end{array} \end{array}$$

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 $T\colon \mathbb{C}^3 \mapsto \mathbb{C}^2$ 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.

$$\begin{aligned}
T(\mathbf{x}+\mathbf{y}) &= T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}+\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}\right)\\
&= T\left(\begin{bmatrix}x_1+y_1\\x_2+y_2\\x_3+y_3\end{bmatrix}\right)\\
&= \begin{bmatrix}2(x_1+y_1)+(x_3+y_3)\\-4(x_2+y_2)\end{bmatrix}\\
&= \begin{bmatrix}(2x_1+x_3)+(2y_1+y_3)\\-4x_2+(-4)y_2\end{bmatrix}\\
&= \begin{bmatrix}2x_1+x_3\\-4x_2\end{bmatrix}+\begin{bmatrix}2y_1+y_3\\-4y_2\end{bmatrix}\\
&= T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right)+T\left(\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}\right)\\
&= T(\mathbf{x})+T(\mathbf{y})
\end{aligned}$$

and

$$\begin{aligned}
T(\alpha\mathbf{x}) &= T\left(\alpha\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right)\\
&= T\left(\begin{bmatrix}\alpha x_1\\\alpha x_2\\\alpha x_3\end{bmatrix}\right)\\
&= \begin{bmatrix}2(\alpha x_1)+(\alpha x_3)\\-4(\alpha x_2)\end{bmatrix}\\
&= \begin{bmatrix}\alpha(2x_1+x_3)\\\alpha(-4x_2)\end{bmatrix}\\
&= \alpha\begin{bmatrix}2x_1+x_3\\-4x_2\end{bmatrix}\\
&= \alpha T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right)\\
&= \alpha T(\mathbf{x})
\end{aligned}$$

So by Definition LT [379], T is a linear transformation. ⊚

It can be just as instructive to look at functions that are *not* linear transformations. Since 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 $S\colon \mathbb{C}^3 \mapsto \mathbb{C}^3$ 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

$$3\,S\left(\begin{bmatrix}1\\2\\3\end{bmatrix}\right) = 3\begin{bmatrix}8\\0\\8\end{bmatrix} = \begin{bmatrix}24\\0\\24\end{bmatrix}$$

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 [379], 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 $T\colon P_3 \mapsto M_{22}$ by

$$T\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix}a + b & a - 2c\\ d & b - d\end{bmatrix}$$

$$\begin{aligned}
T(\mathbf{x} + \mathbf{y}) &= T\left((a_1 + b_1x + c_1x^2 + d_1x^3) + (a_2 + b_2x + c_2x^2 + d_2x^3)\right)\\
&= T\left((a_1 + a_2) + (b_1 + b_2)x + (c_1 + c_2)x^2 + (d_1 + d_2)x^3\right)\\
&= \begin{bmatrix}(a_1 + a_2) + (b_1 + b_2) & (a_1 + a_2) - 2(c_1 + c_2)\\ d_1 + d_2 & (b_1 + b_2) - (d_1 + d_2)\end{bmatrix}\\
&= \begin{bmatrix}(a_1 + b_1) + (a_2 + b_2) & (a_1 - 2c_1) + (a_2 - 2c_2)\\ d_1 + d_2 & (b_1 - d_1) + (b_2 - d_2)\end{bmatrix}\\
&= \begin{bmatrix}a_1 + b_1 & a_1 - 2c_1\\ d_1 & b_1 - d_1\end{bmatrix} + \begin{bmatrix}a_2 + b_2 & a_2 - 2c_2\\ d_2 & b_2 - d_2\end{bmatrix}\\
&= T\left(a_1 + b_1x + c_1x^2 + d_1x^3\right) + T\left(a_2 + b_2x + c_2x^2 + d_2x^3\right)\\
&= T(\mathbf{x}) + T(\mathbf{y})
\end{aligned}$$

and

$$\begin{aligned}
T(\alpha\mathbf{x}) &= T\left(\alpha(a+bx+cx^2+dx^3)\right)\\
&= T\left((\alpha a)+(\alpha b)x+(\alpha c)x^2+(\alpha d)x^3\right)\\
&= \begin{bmatrix}(\alpha a)+(\alpha b) & (\alpha a)-2(\alpha c)\\ \alpha d & (\alpha b)-(\alpha d)\end{bmatrix}\\
&= \begin{bmatrix}\alpha(a+b) & \alpha(a-2c)\\ \alpha d & \alpha(b-d)\end{bmatrix}\\
&= \alpha\begin{bmatrix}a+b & a-2c\\ d & b-d\end{bmatrix}\\
&= \alpha T\left(a+bx+cx^2+dx^3\right)\\
&= \alpha T(\mathbf{x})
\end{aligned}$$

So by Definition LT [379], T is a linear transformation. ⊚

Example LTPP
Linear transformation, polynomials to polynomials
Define a function $S\colon P_4 \mapsto P_5$ by

$$S(p(x)) = (x-2)p(x)$$

Then

$$\begin{aligned}
S(p(x)+q(x)) &= (x-2)(p(x)+q(x)) = (x-2)p(x)+(x-2)q(x) = S(p(x))+S(q(x))\\
S(\alpha p(x)) &= (x-2)(\alpha p(x)) = (x-2)\alpha p(x) = \alpha(x-2)p(x) = \alpha S(p(x))
\end{aligned}$$

So by Definition LT [379], 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 $T\colon U \mapsto V$ 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 throughout this proof to highlight the distinction. Think about your objects.

$$\begin{aligned}
T(\mathbf{0}_U) &= T(\mathbf{0}_U+\mathbf{0}_U) && \text{Zero vector in } U\\
T(\mathbf{0}_U)+\mathbf{0}_V &= T(\mathbf{0}_U+\mathbf{0}_U) && \text{Zero vector in } V\\
T(\mathbf{0}_U)+\mathbf{0}_V &= T(\mathbf{0}_U)+T(\mathbf{0}_U) && \text{Definition LT [379]}\\
\mathbf{0}_V &= T(\mathbf{0}_U) && \text{Theorem VAC [249] in } V
\end{aligned}$$

■

Return to Example NLT [382] 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 [382] and compute $S\left(0 + 0x + 0x^2 + 0x^3\right) = \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$ as an example of Theorem LTTZZ [383] at work.

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 $P\colon \mathbb{C}^4 \mapsto \mathbb{C}^3$ by

$$P\left(\mathbf{x}\right) = A\mathbf{x}$$

So we are using an old friend, the matrix-vector product (Definition MVP [195]) as a way to convert a vector with 4 components into a vector with 3 components. Applying Definition MVP [195] allows us to write the defining formula for P in a slightly different form,

$$P\left(\mathbf{x}\right) = A\mathbf{x} = \begin{bmatrix}3 & -1 & 8 & 1\\2 & 0 & 5 & -2\\1 & 1 & 3 & -7\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = x_1\begin{bmatrix}3\\2\\1\end{bmatrix} + x_2\begin{bmatrix}-1\\0\\1\end{bmatrix} + x_3\begin{bmatrix}8\\5\\3\end{bmatrix} + x_4\begin{bmatrix}1\\-2\\-7\end{bmatrix}$$

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 $\mathbb{C}^3$, Definition CVSM [92] and Definition CVA [91].

$$\begin{aligned}
P(\mathbf{x}) &= A\mathbf{x}\\
&= x_1\begin{bmatrix}3\\2\\1\end{bmatrix} + x_2\begin{bmatrix}-1\\0\\1\end{bmatrix} + x_3\begin{bmatrix}8\\5\\3\end{bmatrix} + x_4\begin{bmatrix}1\\-2\\-7\end{bmatrix}\\
&= \begin{bmatrix}3x_1\\2x_1\\x_1\end{bmatrix} + \begin{bmatrix}-x_2\\0\\x_2\end{bmatrix} + \begin{bmatrix}8x_3\\5x_3\\3x_3\end{bmatrix} + \begin{bmatrix}x_4\\-2x_4\\-7x_4\end{bmatrix}\\
&= \begin{bmatrix}3x_1 - x_2 + 8x_3 + x_4\\2x_1 + 5x_3 - 2x_4\\x_1 + x_2 + 3x_3 - 7x_4\end{bmatrix}
\end{aligned}$$

You might recognize this final expression as being similar in style to some previous examples (Example ALT [380]) and some linear transformations defined in the archetypes. But the expression expressing the output as 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 our matrix operations from Section MO [159], specifically distributivity properties contained in Theorem VSPM [161].

$$\begin{aligned}
P(\mathbf{x}+\mathbf{y}) &= A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = P(\mathbf{x}) + P(\mathbf{y})\\
P(\alpha\mathbf{x}) &= A(\alpha\mathbf{x}) = \alpha(A\mathbf{x}) = \alpha P(\mathbf{x})
\end{aligned}$$

So by Definition LT [379], 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.

Theorem MBLT
Matrices Build Linear Transformations
Suppose that A is an $m \times n$ matrix. Define a function $T\colon \mathbb{C}^n \mapsto \mathbb{C}^m$ by $T(\mathbf{x}) = A\mathbf{x}$. Then T is a linear transformation. □

Proof

$$\begin{aligned}
P(\mathbf{x}+\mathbf{y}) &= A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = P(\mathbf{x}) + P(\mathbf{y})\\
P(\alpha\mathbf{x}) &= A(\alpha\mathbf{x}) = \alpha(A\mathbf{x}) = \alpha P(\mathbf{x})
\end{aligned}$$

So by Definition LT [379], T is a linear transformation. ■

So Theorem MBLT [385] 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 [385] tells that T is a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$, without any further checking.

We can turn Theorem MBLT [385] around. You give me a linear transformation and I will give you a matrix.

Example MFLT
Matrix from a linear transformation
Define the function $R\colon \mathbb{C}^3 \mapsto \mathbb{C}^4$ 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.

$$\begin{aligned}
R\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) &= \begin{bmatrix}2x_1 - 3x_2 + 4x_3\\ x_1 + x_2 + x_3\\ -x_1 + 5x_2 - 3x_3\\ x_2 - 4x_3\end{bmatrix}\\
&= \begin{bmatrix}2x_1\\x_1\\-x_1\\0\end{bmatrix} + \begin{bmatrix}-3x_2\\x_2\\5x_2\\x_2\end{bmatrix} + \begin{bmatrix}4x_3\\x_3\\-3x_3\\-4x_3\end{bmatrix} && \text{Definition CVA [91]}\\
&= x_1\begin{bmatrix}2\\1\\-1\\0\end{bmatrix} + x_2\begin{bmatrix}-3\\1\\5\\1\end{bmatrix} + x_3\begin{bmatrix}4\\1\\-3\\-4\end{bmatrix} && \text{Definition CVSM [92]}\\
&= \begin{bmatrix}2 & -3 & 4\\1 & 1 & 1\\-1 & 5 & -3\\0 & 1 & -4\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} && \text{Definition MVP [195]}
\end{aligned}$$

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 [385], we can easily recognize R as a linear transformation since it has the form described in the hypothesis of the theorem. ⊚

Example MFLT [386] was not accident. Consider any one of the archetypes where both the domain and codomain are sets of column vectors ($\mathbb{C}^m$) 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 $T\colon \mathbb{C}^n \mapsto \mathbb{C}^m$ 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, 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 [218]). 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

$$\mathbf{A}_i = T(\mathbf{e}_i)$$

Then package up these vectors as the columns of a matrix

$$A = [\mathbf{A}_1|\mathbf{A}_2|\mathbf{A}_3|\ldots|\mathbf{A}_n]$$

Does A have the desired properties? First, it is an $m \times n$ matrix. Then

$$\begin{aligned}
T(\mathbf{x}) &= T(I_n\mathbf{x}) && \text{Theorem MMIM [201]}\\
&= T\left([\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\ldots|\mathbf{e}_n]\begin{bmatrix}x_1\\x_2\\x_3\\\vdots\\x_n\end{bmatrix}\right) && \text{Definition SUV [218]}\\
&= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 + \cdots + x_n\mathbf{e}_n) && \text{Definition MVP [195]}\\
&= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + T(x_3\mathbf{e}_3) + \cdots + T(x_n\mathbf{e}_n) && \text{Definition LT [379]}\\
&= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + x_3T(\mathbf{e}_3) + \cdots + x_nT(\mathbf{e}_n) && \text{Definition LT [379]}\\
&= x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + x_3\mathbf{A}_3 + \cdots + x_n\mathbf{A}_n && \text{Definition of } A_i\\
&= [\mathbf{A}_1|\mathbf{A}_2|\mathbf{A}_3|\ldots|\mathbf{A}_n]\begin{bmatrix}x_1\\x_2\\x_3\\\vdots\\x_n\end{bmatrix} && \text{Definition MVP [195]}\\
&= A\mathbf{x}
\end{aligned}$$

as desired. ■

So in the case of vector spaces of column vectors, every matrix leads to a linear transformation (Theorem MBLT [385]), while every linear transformation leads to a matrix (Theorem MLTCV [386]). So matrices and linear transformations are fundamentally the same. We call the matrix A of Theorem MLTCV [386] 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 to more general linear transformations? Yes, that is what Chapter R [441] is all about. Stay tuned. For now, lets illustrate Theorem MLTCV [386] with an example.

Example MOLT
Matrix of a linear transformation
Suppose $S\colon \mathbb{C}^3 \mapsto \mathbb{C}^4$ 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 [386] guarantees that $S\left(\mathbf{x}\right) = C\mathbf{x}$.

As an illuminating exercise, let $\mathbf{z} = \begin{bmatrix}2\\-3\\3\end{bmatrix}$ and compute $S\left(\mathbf{z}\right)$ two different ways. First, return to the definition of S and evaluate $S\left(\mathbf{z}\right)$ directly. Then do the matrix-vector product $C\mathbf{z}$. In both cases you should obtain the vector $S\left(\mathbf{z}\right) = \begin{bmatrix}27\\2\\39\\-12\end{bmatrix}$. ⊚

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 referenced frequently. We have already passed by one occasion to employ it, in the proof of Theorem MLTCV [386]. 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 $T\colon U \mapsto V$ 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

$$T\left(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t\right) = a_1T\left(\mathbf{u}_1\right) + a_2T\left(\mathbf{u}_2\right) + a_3T\left(\mathbf{u}_3\right) + \cdots + a_tT\left(\mathbf{u}_t\right) \quad \square$$

Proof

$$\begin{aligned} &T\left(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t\right) \\ &\quad = T\left(a_1\mathbf{u}_1\right) + T\left(a_2\mathbf{u}_2\right) + T\left(a_3\mathbf{u}_3\right) + \cdots + T\left(a_t\mathbf{u}_t\right) && \text{Definition LT [379]} \\ &\quad = a_1T\left(\mathbf{u}_1\right) + a_2T\left(\mathbf{u}_2\right) + a_3T\left(\mathbf{u}_3\right) + \cdots + a_tT\left(\mathbf{u}_t\right) && \text{Definition LT [379]} \quad \blacksquare \end{aligned}$$

Our next theorem says, informally, that it is enough to know how a linear transformation behaves for inputs from a basis of the domain, and every other output is described by a linear combination of these values. Again, the theorem and its proof are not remarkable, but the insight that goes along with it is fundamental.

Theorem LTDB
Linear Transformation Defined on a Basis
Suppose that $T\colon U \mapsto V$ is a linear transformation, $B = \{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_n\}$ is a basis for U and $\mathbf{w}$ is a vector from U. Let $a_1,\,a_2,\,a_3,\,\ldots,\,a_n$ be scalars from $\mathbb{C}$ such that

$$\mathbf{w} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_n\mathbf{u}_n$$

Then

$$T\left(\mathbf{w}\right) = a_1T\left(\mathbf{u}_1\right) + a_2T\left(\mathbf{u}_2\right) + a_3T\left(\mathbf{u}_3\right) + \cdots + a_nT\left(\mathbf{u}_n\right) \quad \square$$

Proof For any $\mathbf{w} \in U$, Theorem VRRB [287] says there are (unique) scalars such that $\mathbf{w}$ is a linear combination of the basis vectors in B. The result then follows from a straightforward application of Theorem LTLC [389] to the linear combination. $\blacksquare$

Example LTDB1
Linear transformation defined on a basis
Suppose you are told that $T\colon \mathbb{C}^3 \mapsto \mathbb{C}^2$ is a linear transformation and given the 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 [280]), Theorem LTDB [389] says we can compute any output of T with just this information. For example, consider,

$$\mathbf{w} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} = (2) \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + (-3) \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (1) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

so

$$T(\mathbf{w}) = (2) \begin{bmatrix} 2 \\ 1 \end{bmatrix} + (-3) \begin{bmatrix} -1 \\ 4 \end{bmatrix} + (1) \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 13 \\ -11 \end{bmatrix}$$

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 as a linear transformation, along with its values on a handful of vectors (the basis), are just as sufficient as a formula for computing any value of the function. You might notice some parallels between this example and Example MOLT [388] or Theorem MLTCV [386]. ⊚

Example LTDB2
Linear transformation defined on a basis
Suppose you are told that $R\colon \mathbb{C}^3 \mapsto \mathbb{C}^2$ is a linear transformation and given 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 CNSMB [285]). By Theorem LTDB [389] we can compute any output of R with just this information. 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 Theorem LTDB [389] gives us

$$R(\mathbf{y}) = (3)\begin{bmatrix}5\\-1\end{bmatrix} + (-2)\begin{bmatrix}0\\4\end{bmatrix} + (1)\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}17\\-8\end{bmatrix}$$

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.

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 *define* a linear transformation $S\colon P_3 \mapsto M_{22}$ by the values

$$S\left(3 - 4x + x^2\right) = \begin{bmatrix}1 & -3\\2 & 0\end{bmatrix} \qquad S\left(12 - 13x + x^3\right) = \begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$$

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

$$q(x) = 9 - 6x - 5x^2 + 2x^3 = (-5)(3 - 4x + x^2) + (2)(12 - 13x + x^3)$$

Theorem LTDB [389] 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 every other output 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? ⊚

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 $T\colon U \mapsto V$ is a linear transformation. For each $\mathbf{v}$, define the **pre-image** of $\mathbf{v}$ to be the subset of U given by

$$T^{-1}(\mathbf{v}) = \{\mathbf{u} \in U \mid T(\mathbf{u}) = \mathbf{v}\}$$ △

In other words, $T^{-1}(\mathbf{v})$ is the set of all those vectors in the domain U that get "sent" to the vector $\mathbf{v}$. TODO: All preimages form a partition of U, an equivalence relation is about.

Example SPIAS
Sample pre-images, Archetype S
Archetype S [543] is the linear transformation defined by

$$T\colon \mathbb{C}^3 \mapsto M_{22}, \quad T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \begin{bmatrix} a-b & 2a+2b+c \\ 3a+b+c & -2a-6b-2c \end{bmatrix}$$

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 $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \in T^{-1}(\mathbf{v})$. 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 [159]), 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,

$$\begin{bmatrix} 1 & -1 & 0 & 2 \\ 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 3 \\ -2 & -6 & -2 & 2 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & \frac{1}{4} & \frac{5}{4} \\ 0 & \boxed{1} & \frac{1}{4} & -\frac{3}{4} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We regognize this system as having infinitely many solutions described by the single free variable u_3. Eventually obtaining the vector form of the solutions (Theorem VF-

SLS [104]), we can describe the preimage precisely as,

$$\begin{aligned}
T^{-1}(\mathbf{v}) &= \left\{\mathbf{u} \in \mathbb{C}^3 \,\middle|\, T(\mathbf{u}) = \mathbf{v}\right\} \\
&= \left\{ \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \,\middle|\, u_1 = \frac{5}{4} - \frac{1}{4}u_3,\ u_2 = -\frac{3}{4} - \frac{1}{4}u_3 \right\} \\
&= \left\{ \begin{bmatrix} \frac{5}{4} - \frac{1}{4}u_3 \\ -\frac{3}{4} - \frac{1}{4}u_3 \\ u_3 \end{bmatrix} \,\middle|\, u_3 \in \mathbb{C}^3 \right\} \\
&= \left\{ \begin{bmatrix} \frac{5}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + u_3 \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \\ 1 \end{bmatrix} \,\middle|\, u_3 \in \mathbb{C}^3 \right\} \\
&= \begin{bmatrix} \frac{5}{4} \\ -\frac{3}{4} \\ 0 \end{bmatrix} + \mathcal{S}p\left(\left\{ \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \\ 1 \end{bmatrix} \right\}\right)
\end{aligned}$$

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 [206]. 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 [159]), 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,

$$\begin{bmatrix} 1 & -1 & 0 & 1 \\ 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 2 \\ -2 & -6 & -2 & 4 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} \boxed{1} & 0 & \frac{1}{4} & 0 \\ 0 & \boxed{1} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem RCLS [56] we recognize this system as inconsistent. So no vector $\mathbf{u}$ is a member of $T^{-1}(\mathbf{v})$ and so

$$T^{-1}(\mathbf{v}) = \emptyset$$

⊚

The preimage is just a set, it is rarely a subspace of U (you might think about just when it is a subspace). We will describe its properties going forward, but in some ways it will be a notational convenience as much as anything else.

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 $T\colon U \mapsto V$ and $S\colon U \mapsto V$ are two linear transformations with the same domain and codomain. Then their **sum** is the function $T + S\colon U \mapsto V$ whose outputs are defined by

$$(T + S)(\mathbf{u}) = T(\mathbf{u}) + S(\mathbf{u})$$ △

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 [394] only provides a function. It would be nice to know that if the constituents (T, S) are linear transformations, then so is $T + S$.

Theorem SLTLT
Sum of Linear Transformations is a Linear Transformation
Suppose that $T\colon U \mapsto V$ and $S\colon U \mapsto V$ are two linear transformations with the same domain and codomain. Then $T + S\colon U \mapsto V$ is a linear transformation. □

Proof We simply check the defining properties of a linear transformation (Definition LT [379]). This is a good place to consistently ask yourself which objects are being combined with which operations.

$$\begin{aligned}
(T + S)(\mathbf{x} + \mathbf{y}) &= T(\mathbf{x} + \mathbf{y}) + S(\mathbf{x} + \mathbf{y}) && \text{Definition LTA [394]}\\
&= T(\mathbf{x}) + T(\mathbf{y}) + S(\mathbf{x}) + S(\mathbf{y}) && \text{Definition LT [379]}\\
&= T(\mathbf{x}) + S(\mathbf{x}) + T(\mathbf{y}) + S(\mathbf{y}) && \text{Commutativity in } V\\
&= (T + S)(\mathbf{x}) + (T + S)(\mathbf{y}) && \text{Definition LTA [394]}
\end{aligned}$$

and

$$\begin{aligned}
(T + S)(\alpha\mathbf{x}) &= T(\alpha\mathbf{x}) + S(\alpha\mathbf{x}) && \text{Definition LTA [394]}\\
&= \alpha T(\mathbf{x}) + \alpha S(\mathbf{x}) && \text{Definition LT [379]}\\
&= \alpha(T(\mathbf{x}) + S(\mathbf{x})) && \text{Distributivity in } V\\
&= \alpha(T + S)(\mathbf{x}) && \text{Definition LTA [394]}
\end{aligned}$$

■

Example STLT
Sum of two linear transformations
Suppose that $T\colon \mathbb{C}^2 \mapsto \mathbb{C}^3$ and $S\colon \mathbb{C}^2 \mapsto \mathbb{C}^3$ are defined by

$$T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}x_1+2x_2\\3x_1-4x_2\\5x_1+2x_2\end{bmatrix} \qquad S\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) = \begin{bmatrix}4x_1-x_2\\x_1+3x_2\\-7x_1+5x_2\end{bmatrix}$$

Then by Definition LTA [394], 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 [394] 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 $T\colon U \mapsto V$ is a linear transformation and $\alpha \in \mathbb{C}$. Then the **scalar multiple** is the function $\alpha T\colon U \mapsto V$ whose outputs are defined by

$$(\alpha T)(\mathbf{u}) = \alpha T(\mathbf{u})$$ △

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 $T\colon U \mapsto V$ is a linear transformation and $\alpha \in \mathbb{C}$. Then $(\alpha T)\colon U \mapsto V$ is a linear transformation. □

Proof We simply check the defining properties of a linear transformation (Definition LT [379]). This is another good place to consistently ask yourself which objects are being combined with which operations.

$$\begin{aligned}
(\alpha T)(\mathbf{x}+\mathbf{y}) &= \alpha\left(T(\mathbf{x}+\mathbf{y})\right) && \text{Definition LTSM [395]}\\
&= \alpha\left(T(\mathbf{x})+T(\mathbf{y})\right) && \text{Definition LT [379]}\\
&= \alpha T(\mathbf{x})+\alpha T(\mathbf{y}) && \text{Distributivity in } V\\
&= (\alpha T)(\mathbf{x})+(\alpha T)(\mathbf{y}) && \text{Definition LTSM [395]}
\end{aligned}$$

and

$$\begin{aligned}
(\alpha T)(\beta\mathbf{x}) &= \alpha T(\beta\mathbf{x}) && \text{Definition LTSM [395]}\\
&= \alpha\left(\beta T(\mathbf{x})\right) && \text{Definition LT [379]}\\
&= (\alpha\beta)T(\mathbf{x}) && \text{Associativity in } V\\
&= (\beta\alpha)T(\mathbf{x}) && \text{Commutativity}\\
&= \beta\left(\alpha T(\mathbf{x})\right) && \text{Associativity in } V\\
&= \beta(\alpha T)(\mathbf{x}) && \text{Definition LTSM [395]}
\end{aligned}$$

■

Example SMLT
Scalar multiple of a linear transformation
Suppose that $T\colon \mathbb{C}^4 \mapsto \mathbb{C}^3$ 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 [395], 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 [395] we know $T+S$ is also a linear transformation from $\mathbb{C}^4$ to $\mathbb{C}^3$. ⊚

Now, lets 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 $LT(U,\,V)$. Definition LTA [394] and Definition LTSM [395] tell us how we can "add" and "scalar multiply" two elements of $LT(U,\,V)$. Theorem SLTLT [394] and Theorem MLTLT [395] tell us that if we do these operations, then the resulting functions are linear transformations that are also in $LT(U,\,V)$. 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, $LT(U,\,V)$ is a vector space when the operations are those given in Definition LTA [394] and Definition LTSM [395]. □

Proof Theorem SLTLT [394] and Theorem MLTLT [395] provide two of the ten axioms in Definition VS [239]. However, we still need to verify the remaining eight axioms. By and large, the proofs are straightforward and rely on concocting the obvious object, or by reducing the question to the same vector space axiom 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 $Z\colon U \mapsto V$ defined by $Z(\mathbf{u}) = \mathbf{0}_V$ for every $\mathbf{u} \in U$. Notice how we do not need to know any specifics about U and V to make this definition. ■

Definition LTC
Linear Transformation Composition
Suppose that $T\colon U \mapsto V$ and $S\colon V \mapsto W$ are linear transformations. Then the **composition** of S and T is the function $(S \circ T)\colon U \mapsto W$ whose outputs are defined by

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u})) \qquad \triangle$$

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 $T\colon U \mapsto V$ and $S\colon V \mapsto W$ are linear transformations. Then $(S \circ T)\colon U \mapsto W$ is a linear transformation. □

Proof We simply check the defining properties of a linear transformation (Definition LT [379]).

$$\begin{aligned}
(S \circ T)(\mathbf{x}+\mathbf{y}) &= S(T(\mathbf{x}+\mathbf{y})) && \text{Definition LTC [397]}\\
&= S(T(\mathbf{x})+T(\mathbf{y})) && \text{Definition LT [379] for } T\\
&= S(T(\mathbf{x}))+S(T(\mathbf{y})) && \text{Definition LT [379] for } S\\
&= (S \circ T)(\mathbf{x})+(S \circ T)(\mathbf{y}) && \text{Definition LTC [397]}
\end{aligned}$$

and

$$\begin{aligned}
(S \circ T)(\alpha\mathbf{x}) &= S(T(\alpha\mathbf{x})) && \text{Definition LTC [397]}\\
&= S(\alpha T(\mathbf{x})) && \text{Definition LT [379] for } T\\
&= \alpha S(T(\mathbf{x})) && \text{Definition LT [379] for } S\\
&= \alpha(S \circ T)(\mathbf{x}) && \text{Definition LTC [397]}
\end{aligned}$$ ■

Example CTLT
Composition of two linear transformations
Suppose that $T\colon \mathbb{C}^2 \mapsto \mathbb{C}^4$ and $S\colon \mathbb{C}^4 \mapsto \mathbb{C}^3$ 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 [397]

$$\begin{aligned}
(S \circ T)\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right) &= S\left(T\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}\right)\right)\\
&= S\left(\begin{bmatrix}x_1+2x_2\\3x_1-4x_2\\5x_1+2x_2\\6x_1-3x_2\end{bmatrix}\right)\\
&= \begin{bmatrix}2(x_1+2x_2)-(3x_1-4x_2)+(5x_1+2x_2)-(6x_1-3x_2)\\5(x_1+2x_2)-3(3x_1-4x_2)+8(5x_1+2x_2)-2(6x_1-3x_2)\\-4(x_1+2x_2)+3(3x_1-4x_2)-4(5x_1+2x_2)+5(6x_1-3x_2)\end{bmatrix}\\
&= \begin{bmatrix}-2x_1+13x_2\\24x_1+44x_2\\15x_1-43x_2\end{bmatrix}
\end{aligned}$$

and by Theorem CLTLT [397] $S \circ T$ is a linear transformation from $\mathbb{C}^2$ to $\mathbb{C}^3$. ⊚

Here is an interesting exercise that will presage an important result later. In Example STLT [395] compute (via Theorem MLTCV [386]) the matrix of T, S and $T+S$. Do you see a relationship between these three matrices?

In Example SMLT [396] compute (via Theorem MLTCV [386]) the matrix of T and $2T$. Do you see a relationship between these two matrices?

Here's the tough one. In Example CTLT [397] compute (via Theorem MLTCV [386]) 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?

$$T\colon \mathbb{C}^3 \mapsto \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 - x_2 + x_3 \\ 8x_2 - 6 \end{bmatrix}$$

2. Determine the matrix representation of the linear transformation S below.

$$S\colon \mathbb{C}^2 \mapsto \mathbb{C}^3, \quad S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 + 5x_2 \\ 8x_1 - 3x_2 \\ -4x_1 \end{bmatrix}$$

3. Theorem LTLC [389] has a fairly simple proof. Yet the result itself is very powerful. Comment on why we might say this.

Section ILT
Injective Linear Transformations

Linear transformations have 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 sets like the null space and the range. 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 $T\colon U \mapsto V$ 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 [538] is the linear transformation

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} -2x_1 + 3x_2 + 3x_3 - 6x_4 + 3x_5 \\ -16x_1 + 9x_2 + 12x_3 - 28x_4 + 28x_5 \\ -19x_1 + 7x_2 + 14x_3 - 32x_4 + 37x_5 \\ -21x_1 + 9x_2 + 15x_3 - 35x_4 + 39x_5 \\ -9x_1 + 5x_2 + 7x_3 - 16x_4 + 16x_5 \end{bmatrix}$$

Notice that for

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ -1 \\ 2 \\ 4 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 4 \\ 7 \\ 0 \\ 5 \\ 7 \end{bmatrix}$$

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 [399]. 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. ⊚

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 [399]. 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 [541] is the linear transformation

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} -65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5 \\ 36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5 \\ -44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5 \\ 34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5 \\ 12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5 \end{bmatrix}$$

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,

$$
\begin{aligned}
T(\mathbf{x}) &= T(\mathbf{y}) \\
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) \\
\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} \\
\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} &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \\
\begin{bmatrix} -65(x_1 - y_1) + 128(x_2 - y_2) + 10(x_3 - y_3) - 262(x_4 - y_4) + 40(x_5 - y_5) \\ 36(x_1 - y_1) - 73(x_2 - y_2) - (x_3 - y_3) + 151(x_4 - y_4) - 16(x_5 - y_5) \\ -44(x_1 - y_1) + 88(x_2 - y_2) + 5(x_3 - y_3) - 180(x_4 - y_4) + 24(x_5 - y_5) \\ 34(x_1 - y_1) - 68(x_2 - y_2) - 3(x_3 - y_3) + 140(x_4 - y_4) - 18(x_5 - y_5) \\ 12(x_1 - y_1) - 24(x_2 - y_2) - (x_3 - y_3) + 49(x_4 - y_4) - 5(x_5 - y_5) \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \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} x_1 - y_1 \\ x_2 - y_2 \\ x_3 - y_3 \\ x_4 - y_4 \\ x_5 - y_5 \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}
\end{aligned}
$$

Now we recognize that we have a homogenous system of 5 equations in 5 variables (the terms $x_i - y_i$ are the variables), so we row-reduce the coefficient matrix to

$$
\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}
$$

So the only solution is the trivial solution

$$
x_1 - y_1 = 0 \qquad x_2 - y_2 = 0 \qquad x_3 - y_3 = 0 \qquad x_4 - y_4 = 0 \qquad x_5 - y_5 = 0
$$

and we conlude that indeed $\mathbf{x} = \mathbf{y}$. By Definition ILT [399], T is injective. ⊚

Lets now examine an injective linear transformation between abstract vector spaces.

Example IAV
Injective, Archetype V
Archetype V [544] is defined by

$$T\colon P_3 \mapsto M_{22},\quad T\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

To establish that the linear transformation is injective, begin by supposing that two polynomial inputs yield the same output matrix,

$$T\left(a_1 + b_1x + c_1x^2 + d_1x^3\right) = T\left(a_2 + b_2x + c_2x^2 + d_2x^3\right)$$

Then

$$\begin{aligned}
\mathcal{O} &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\
&= T\left(a_1 + b_1x + c_1x^2 + d_1x^3\right) - T\left(a_2 + b_2x + c_2x^2 + d_2x^3\right) && \text{Hypothesis} \\
&= T\left((a_1 + b_1x + c_1x^2 + d_1x^3) - (a_2 + b_2x + c_2x^2 + d_2x^3)\right) && T \text{ linear transformation} \\
&= T\left((a_1 - a_2) + (b_1 - b_2)x + (c_1 - c_2)x^2 + (d_1 - d_2)x^3\right) && \text{Operations in } P_3 \\
&= \begin{bmatrix} (a_1 - a_2) + (b_1 - b_2) & (a_1 - a_2) - 2(c_1 - c_2) \\ (d_1 - d_2) & (b_1 - b_2) - (d_1 - d_2) \end{bmatrix} && \text{Definition of } T
\end{aligned}$$

This single matrix equality translates to the homogenous system of equations in the variables $a_i - b_i$,

$$\begin{aligned}
(a_1 - a_2) + (b_1 - b_2) &= 0 \\
(a_1 - a_2) - 2(c_1 - c_2) &= 0 \\
(d_1 - d_2) &= 0 \\
(b_1 - b_2) - (d_1 - d_2) &= 0
\end{aligned}$$

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.

$$a_1 - a_2 = 0 \qquad b_1 - b_2 = 0 \qquad c_1 - c_2 = 0 \qquad d_1 - d_2 = 0$$

so that

$$a_1 = a_2 \qquad b_1 = b_2 \qquad c_1 = c_2 \qquad d_1 = d_2$$

so the two inputs must be equal polynomials. By Definition ILT [399], T is injective. ⊚

Subsection NSLT
Null Space of a Linear Transformation

For a linear transformation $T\colon U \mapsto V$, the null space 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 NSLT
Null Space of a Linear Transformation
Suppose $T\colon U \mapsto V$ is a linear transformation. Then the **null space** of T is the set $\mathcal{N}(T) = \{\mathbf{u} \in U \mid T(\mathbf{u}) = \mathbf{0}\}$ △

Example NNSAO
Nontrivial null space, Archetype O
Archetype O [533] is the linear transformation

$$T\colon \mathbb{C}^3 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -x_1 + x_2 - 3x_3 \\ -x_1 + 2x_2 - 4x_3 \\ x_1 + x_2 + x_3 \\ 2x_1 + 3x_2 + x_3 \\ x_1 + 2x_3 \end{bmatrix}$$

To determine the elements of $\mathbb{C}^3$ in $\mathcal{N}(T)$, find those vectors $\mathbf{u}$ such that $T(\mathbf{u}) = \mathbf{0}$, that is,

$$T(\mathbf{u}) = \mathbf{0}$$
$$\begin{bmatrix} -u_1 + u_2 - 3u_3 \\ -u_1 + 2u_2 - 4u_3 \\ u_1 + u_2 + u_3 \\ 2u_1 + 3u_2 + u_3 \\ u_1 + 2u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Vector equality (Definition CVE [90]) leads us to a homogenous system of 5 equations in the variables u_i,

$$\begin{aligned} -u_1 + u_2 - 3u_3 &= 0 \\ -u_1 + 2u_2 - 4u_3 &= 0 \\ u_1 + u_2 + u_3 &= 0 \\ 2u_1 + 3u_2 + u_3 &= 0 \\ u_1 + 2u_3 &= 0 \end{aligned}$$

Row-reducing the coefficient matrix gives

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

The null space of T is the set of solutions to this homogenous system of equations, which by Theorem BNS [138] can be expressed as

$$\mathcal{N}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$ ⊚

We know that the span of a set of vectors is always a subspace (Theorem SSS [260]), so the null space computed in Example NNSAO [403] is also a subspace. This is no accident, the null space of a linear transformation is *always* a subspace.

Theorem NSLTS
Null Space of a Linear Transformation is a Subspace
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then the null space of T, $\mathcal{N}(T)$, is a subspace of U. □

Proof We can apply the three-part test of Theorem TSS [255]. First $T(\mathbf{0}_U) = \mathbf{0}_V$ by Theorem LTTZZ [383], so $\mathbf{0}_U \in \mathcal{N}(T)$ and we know that the null space is non-empty.

Suppose we assume that $\mathbf{x}, \mathbf{y} \in \mathcal{N}(T)$. Is $\mathbf{x} + \mathbf{y} \in \mathcal{N}(T)$?

$$\begin{aligned} T(\mathbf{x}+\mathbf{y}) &= T(\mathbf{x}) + T(\mathbf{y}) && \text{Definition LT [379]} \\ &= \mathbf{0} + \mathbf{0} && \mathbf{x}, \mathbf{y} \in \mathcal{N}(T) \\ &= \mathbf{0} && \text{Property of zero vector} \end{aligned}$$

This qualifies $\mathbf{x} + \mathbf{y}$ for membership in $\mathcal{N}(T)$. So we have additive closure.

Suppose we assume that $\alpha \in \mathbb{C}$ and $\mathbf{x} \in \mathcal{N}(T)$. Is $\alpha\mathbf{x} \in \mathcal{N}(T)$?

$$\begin{aligned} T(\alpha\mathbf{x}) &= \alpha T(\mathbf{x}) && \text{Definition LT [379]} \\ &= \alpha\mathbf{0} && \mathbf{x} \in \mathcal{N}(T) \\ &= \mathbf{0} && \text{Theorem ZVSM [247]} \end{aligned}$$ ■

This qualifies $\alpha\mathbf{x}$ for membership in $\mathcal{N}(T)$. So we have scalar closure and Theorem TSS [255] tells us that $\mathcal{N}(T)$ is a subspace of U.

Let's compute another null space, now that we know in advance that it will be a subspace.

Example TNSAP
Trivial null space, Archetype P
Archetype P [536] is the linear transformation

$$T\colon \mathbb{C}^3 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -x_1 + x_2 + x_3 \\ -x_1 + 2x_2 + 2x_3 \\ x_1 + x_2 + 3x_3 \\ 2x_1 + 3x_2 + x_3 \\ -2x_1 + x_2 + 3x_3 \end{bmatrix}$$

To determine the elements of $\mathbb{C}^3$ in $\mathcal{N}(T)$, find those vectors $\mathbf{u}$ such that $T(\mathbf{u}) = \mathbf{0}$, that is,

$$T(\mathbf{u}) = \mathbf{0}$$
$$\begin{bmatrix} -u_1 + u_2 + u_3 \\ -u_1 + 2u_2 + 2u_3 \\ u_1 + u_2 + 3u_3 \\ 2u_1 + 3u_2 + u_3 \\ -2u_1 + u_2 + 3u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Vector equality (Definition CVE [90]) leads us to a homogeneous system of 5 equations in the variables u_i,

$$\begin{aligned} -u_1 + u_2 + u_3 &= 0 \\ -u_1 + 2u_2 + 2u_3 &= 0 \\ u_1 + u_2 + 3u_3 &= 0 \\ 2u_1 + 3u_2 + u_3 &= 0 \\ -2u_1 + u_2 + 3u_3 &= 0 \end{aligned}$$

Row-reducing the coefficient matrix gives

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The null space of T is the set of solutions to this homogeneous system of equations, which is simply the trivial solution $\mathbf{u} = \mathbf{0}$, so

$$\mathcal{N}(T) = \{\mathbf{0}\} = \mathcal{S}p(\{\ \})$$ ⊚

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 null space.

Theorem NSPI
Null Space and Pre-Image
Suppose $T\colon U \mapsto V$ 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

$$T^{-1}(\mathbf{v}) = \{\mathbf{u} + \mathbf{z} \mid \mathbf{z} \in \mathcal{N}(T)\} = \mathbf{u} + \mathcal{N}(T)$$ □

Proof Let $M = \{\mathbf{u} + \mathbf{z} \mid \mathbf{z} \in \mathcal{N}(T)\}$. 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 $\mathbf{z} \in \mathcal{N}(T)$. Then

$$\begin{aligned} T(\mathbf{w}) &= T(\mathbf{u} + \mathbf{z}) \\ &= T(\mathbf{u}) + T(\mathbf{z}) && \text{Definition LT [379]} \\ &= \mathbf{v} + \mathbf{0} && \mathbf{u} \in T^{-1}(\mathbf{v}),\ \mathbf{z} \in \mathcal{N}(T) \\ &= \mathbf{v} && \text{Property of } \mathbf{0} \end{aligned}$$

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,

$$\begin{aligned} T(\mathbf{x} - \mathbf{u}) &= T(\mathbf{x}) - T(\mathbf{u}) && \text{Definition LT [379]} \\ &= \mathbf{v} - \mathbf{v} && \mathbf{x}, \mathbf{u} \in T^{-1}(\mathbf{v}) \\ &= \mathbf{0} \end{aligned}$$

This qualifies $\mathbf{x} - \mathbf{u}$ for membership in the null space of T, $\mathcal{N}(T)$. So there is a vector $\mathbf{z} \in \mathcal{N}(T)$ 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 [206]. Additionally, you might go back and review Example SPIAS [392]. Can you tell now which is the only preimage to be a subspace?

Theorem NSILT
Null Space of an Injective Linear Transformation
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then T is injective if and only if the null space of T is trivial, $\mathcal{N}(T) = \{\mathbf{0}\}$. □

Proof ($\Rightarrow$) Suppose $\mathbf{x} \in \mathcal{N}(T)$. Then by Definition NSLT [403], $T(\mathbf{x}) = \mathbf{0}$. By Theorem LTTZZ [383], $T(\mathbf{0}) = \mathbf{0}$. Now, since $T(\mathbf{x}) = T(\mathbf{0})$, Definition ILT [399] we conclude that $\mathbf{x} = \mathbf{0}$. Therefore $\mathcal{N}(T) = \{\mathbf{0}\}$.

($\Leftarrow$) To establish that T is injective, appeal to Definition ILT [399] and begin with the assumption that $T(\mathbf{x}) = T(\mathbf{y})$. Then

$$\begin{aligned} \mathbf{0} &= T(\mathbf{x}) - T(\mathbf{y}) && \text{Hypothesis} \\ &= T(\mathbf{x} - \mathbf{y}) && \text{Definition LT [379]} \end{aligned}$$

so by Definition NSLT [403] and the hypothesis that the null space is trivial,

$$\mathbf{x} - \mathbf{y} \in \mathcal{N}(T) = \{\mathbf{0}\}$$

which means that

$$\begin{aligned} \mathbf{0} &= \mathbf{x} - \mathbf{y} \\ \mathbf{x} &= \mathbf{y} \end{aligned}$$

thus establishing that T is injective. ■

Example NIAQR
Not injective, Archetype Q, revisited
We are now in a position to revisit our first example in this section, Example NIAQ [399]. In that example, we showed that Archetype Q [538] is not injective by constructing two vectors, which when used to evaluate the linear transformation provided the same output, thus violating Definition ILT [399]. 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 $\mathcal{N}(T)$ for Archetype Q [538]. Choose a vector $\mathbf{x}$ at random, and then compute $\mathbf{y} = \mathbf{x} + \mathbf{z}$ (verify this computation back in Example NIAQ [399]). Then

$$\begin{aligned}
T(\mathbf{y}) &= T(\mathbf{x} + \mathbf{z}) \\
&= T(\mathbf{x}) + T(\mathbf{z}) && \text{Definition LT [379]} \\
&= T(\mathbf{x}) + \mathbf{0} && \mathbf{z} \in \mathcal{N}(T) \\
&= T(\mathbf{x}) && \text{Property of } \mathbf{0}
\end{aligned}$$

Whenever the null space 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 NSILT [406]. For an injective linear transformation, the null space is trivial and our only choice for $\mathbf{z}$ is the zero vector, which will not help us create two *different* inputs for T that 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 NNSAO [403] the null space of Archetype O [533] was determined to be

$$\mathcal{S}p\left(\left\{\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$

a subspace of $\mathbb{C}^3$ with dimension 1. Since the null space is not trivial, Theorem NSILT [406] tells us that T is not injective. ⊚

Example IAP
Injective, Archetype P
In Example TNSAP [405] it was shown that the linear transformation in Archetype P [536] has a trivial null space. So by Theorem NSILT [406], T is injective. ⊚

There is a connection between injective linear transformations and linear independent sets that we will make precise in the next two theorems. However, more informally, we can get a 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 if the trivial way, both input vectors are equal.

Theorem ILTLI
Injective Linear Transformations and Linear Independence
Suppose that $T\colon U \mapsto V$ 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 S (Definition RLD [271], Definition LI [271]),

$$\begin{aligned}
a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) + a_3T(\mathbf{u}_3) + \ldots + a_tT(\mathbf{u}_t) &= \mathbf{0} && \\
T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t) &= \mathbf{0} && \text{Theorem LTLC [389]}\\
a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t &\in \mathcal{N}(T) && \text{Definition NSLT [403]}\\
a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t &\in \{\mathbf{0}\} && \text{Theorem NSILT [406]}\\
a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t &= \mathbf{0} &&
\end{aligned}$$

Since this is a relation of linear dependence on the linearly independent set S, we can conclude that

$$a_1 = 0 \qquad a_2 = 0 \qquad a_3 = 0 \qquad \ldots \qquad a_t = 0$$

and this establishes that R is a linearly independent set. ■

Theorem ILTB
Injective Linear Transformations and Bases
Suppose that $T\colon U \mapsto V$ 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 injective if and only if $C = \{T(\mathbf{u}_1),\,T(\mathbf{u}_2),\,T(\mathbf{u}_3),\,\ldots,\,T(\mathbf{u}_m)\}$ is a linearly independent subset of V. □

Proof (⇒) Assume T is injective. Since B is a basis, we know B is linearly independent (Definition B [279]). Then Theorem ILTLI [408] says that C is a linearly independent subset of V.

(⇐) Assume that C is linearly independent. To establish that T is injective, we will show that the null space of T is trivial (Theorem NSILT [406]). Suppose that $\mathbf{u} \in \mathcal{N}(T)$. As an element of U, we can write $\mathbf{u}$ as a linear combination of the basis vectors in B (uniquely). So there are are scalars, $a_1,\,a_2,\,a_3,\,\ldots,\,a_m$, such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_m\mathbf{u}_m$$

Then,

$$\begin{aligned}
\mathbf{0} &= T(\mathbf{u}) && \mathbf{u} \in \mathcal{N}(T)\\
&= T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_m\mathbf{u}_m) && B \text{ spans } U\\
&= a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) + a_3T(\mathbf{u}_3) + \cdots + a_mT(\mathbf{u}_m) && \text{Theorem LTLC [389]}
\end{aligned}$$

This is a relation of linear dependence (Definition RLD [271]) on the linearly independent set C, so the scalars are all zero: $a_1 = a_2 = a_3 = \cdots = a_m = 0$. Then

$$\begin{aligned}
\mathbf{u} &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_m\mathbf{u}_m \\
&= 0\mathbf{u}_1 + 0\mathbf{u}_2 + 0\mathbf{u}_3 + \cdots + 0\mathbf{u}_m \\
&= \mathbf{0} + \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} && \text{Theorem ZSSM [247]} \\
&= \mathbf{0} && \text{Property of } \mathbf{0}
\end{aligned}$$

Since $\mathbf{u}$ was chosen as an arbitrary vector from $\mathcal{N}(T)$, we have $\mathcal{N}(T) = \{\mathbf{0}\}$ and Theorem NSILT [406] tells us that T is injective. ■

Subsection ILTD
Injective Linear Transformations and Dimension

Theorem ILTD
Injective Linear Transformations and Dimension
Suppose that $T\colon U \mapsto V$ is an injective linear transformation. Then $\dim(U) \leq \dim(V)$.□

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 [408], C is linearly independent and therefore must contain m distinct vectors. So we have a set of m linearly independent vectors in V, a vector space of dimension t, with $m > t$. However, this contradicts Theorem G [310], 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 [543] is

$$T\colon M_{23} \mapsto \mathbb{C}^4, \quad T\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a + 2b + 12c - 3d + e + 6f \\ 2a - b - c + d - 11f \\ a + b + 7c + 2d + e - 3f \\ a + 2b + 12c + 5e - 5f \end{bmatrix}$$

Since $\dim(M_{23}) = 6 > 4 = \dim(\mathbb{C}^4)$, T cannot be injective for then it would violate Theorem ILTD [409]. ⊚

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 [529] and Archetype N [531] 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 [394] we saw how to combine linear transformations to build new linear transformations, specifically, how to build the composition of two linear transformations (Definition LTC [397]). 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 $T\colon U \mapsto V$ and $S\colon V \mapsto W$ are injective linear transformations. Then $(S \circ T)\colon U \mapsto W$ is an injective linear transformation. □

Proof That the composition is a linear transformation was established in Theorem CLTLT [397], so we need only establish that the composition is injective. Applying Definition ILT [399], choose $\mathbf{x}$, $\mathbf{y}$ from U. Then

$$\begin{aligned}
(S \circ T)(\mathbf{x}) &= (S \circ T)(\mathbf{y}) \\
S(T(\mathbf{x})) &= S(T(\mathbf{y})) && \text{Definition LTC [397]} \\
T(\mathbf{x}) &= T(\mathbf{y}) && S \text{ is injective} \\
\mathbf{x} &= \mathbf{y} && T \text{ is injective}
\end{aligned}$$

■

Subsection READ
Reading Questions

1. Suppose $T\colon \mathbb{C}^8 \mapsto \mathbb{C}^5$ is a linear transformation. Why can't T be injective?

2. Describe the null space of a injective linear transformation.

3. Theorem NSPI [406] should remind you of Theorem PSPHS [206]. Why do we say this?

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 [399] and note the parallels and the contrasts. In the next section, Section IVLT [425], we will combine the two properties.

As usual, we lead with a definition.

Definition SLT
Surjective Linear Transformation
Suppose $T\colon U \mapsto V$ 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 [538] is the linear transformation

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} -2x_1 + 3x_2 + 3x_3 - 6x_4 + 3x_5 \\ -16x_1 + 9x_2 + 12x_3 - 28x_4 + 28x_5 \\ -19x_1 + 7x_2 + 14x_3 - 32x_4 + 37x_5 \\ -21x_1 + 9x_2 + 15x_3 - 35x_4 + 39x_5 \\ -9x_1 + 5x_2 + 7x_3 - 16x_4 + 16x_5 \end{bmatrix}$$

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

$$\begin{aligned}
\begin{bmatrix} -1 \\ 2 \\ 3 \\ -1 \\ 4 \end{bmatrix} = \mathbf{v} = T(\mathbf{u}) &= T\left(\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}\right) \\
&= \begin{bmatrix} -2u_1 + 3u_2 + 3u_3 - 6u_4 + 3u_5 \\ -16u_1 + 9u_2 + 12u_3 - 28u_4 + 28u_5 \\ -19u_1 + 7u_2 + 14u_3 - 32u_4 + 37u_5 \\ -21u_1 + 9u_2 + 15u_3 - 35u_4 + 39u_5 \\ -9u_1 + 5u_2 + 7u_3 - 16u_4 + 16u_5 \end{bmatrix} \\
&= \begin{bmatrix} -2 & 3 & 3 & -6 & 3 \\ -16 & 9 & 12 & -28 & 28 \\ -19 & 7 & 14 & -32 & 37 \\ -21 & 9 & 15 & -35 & 39 \\ -9 & 5 & 7 & -16 & 16 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}
\end{aligned}$$

Now we recognize the appropriate input vector $\mathbf{u}$ as a solution to a linear system of equations. Form the augmented matrix, and row-reduce to

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

With a leading 1 in the last column, Theorem RCLS [56] tells us the system is inconsistent. From this we conclude that no such vector $\mathbf{u}$ exists, and by Definition SLT [411], 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 [411]. 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 [541] is the linear transformation

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5 \end{bmatrix}\right) = \begin{bmatrix} -65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5\\ 36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5\\ -44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5\\ 34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5\\ 12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5 \end{bmatrix}$$

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,

$$\begin{aligned} T(\mathbf{u}) &= \mathbf{v}\\ \begin{bmatrix} -65u_1 + 128u_2 + 10u_3 - 262u_4 + 40u_5\\ 36u_1 - 73u_2 - u_3 + 151u_4 - 16u_5\\ -44u_1 + 88u_2 + 5u_3 - 180u_4 + 24u_5\\ 34u_1 - 68u_2 - 3u_3 + 140u_4 - 18u_5\\ 12u_1 - 24u_2 - u_3 + 49u_4 - 5u_5 \end{bmatrix} &= \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\\ v_5 \end{bmatrix}\\ \begin{bmatrix} -65 & 128 & 10 & -262 & 40\\ 36 & -73 & -1 & 151 & -16\\ -44 & 88 & 5 & -180 & 24\\ 34 & -68 & -3 & 140 & -18\\ 12 & -24 & -1 & 49 & -5 \end{bmatrix} \begin{bmatrix} u_1\\ u_2\\ u_3\\ u_4\\ u_5 \end{bmatrix} &= \begin{bmatrix} v_1\\ v_2\\ v_3\\ v_4\\ v_5 \end{bmatrix} \end{aligned}$$

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 NSI [228], Definition MI [216]).

$$\begin{bmatrix} -65 & 128 & 10 & -262 & 40\\ 36 & -73 & -1 & 151 & -16\\ -44 & 88 & 5 & -180 & 24\\ 34 & -68 & -3 & 140 & -18\\ 12 & -24 & -1 & 49 & -5 \end{bmatrix}^{-1} = \begin{bmatrix} -47 & 92 & 1 & -181 & -14\\ 27 & -55 & \frac{7}{2} & \frac{221}{4} & 11\\ -32 & 64 & -1 & -126 & -12\\ 25 & -50 & \frac{3}{2} & \frac{199}{2} & 9\\ 9 & -18 & \frac{1}{2} & \frac{71}{2} & 4 \end{bmatrix}$$

so we find that

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} = \begin{bmatrix} -47 & 92 & 1 & -181 & -14 \\ 27 & -55 & \frac{7}{2} & \frac{221}{4} & 11 \\ -32 & 64 & -1 & -126 & -12 \\ 25 & -50 & \frac{3}{2} & \frac{199}{2} & 9 \\ 9 & -18 & \frac{1}{2} & \frac{71}{2} & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix}$$

$$= \begin{bmatrix} -47v_1 + 92v_2 + v_3 - 181v_4 - 14v_5 \\ 27v_1 - 55v_2 + \frac{7}{2}v_3 + \frac{221}{4}v_4 + 11v_5 \\ -32v_1 + 64v_2 - v_3 - 126v_4 - 12v_5 \\ 25v_1 - 50v_2 + \frac{3}{2}v_3 + \frac{199}{2}v_4 + 9v_5 \\ 9v_1 - 18v_2 + \frac{1}{2}v_3 + \frac{71}{2}v_4 + 4v_5 \end{bmatrix}$$

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 [411] 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 $\mathbf{v}$. ⊚

Lets now examine a surjective linear transformation between abstract vector spaces.

Example SAV
Surjective, Archetype V
Archetype V [544] is defined by

$$T\colon P_3 \mapsto M_{22}, \quad T\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

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,

$$\begin{aligned} \begin{bmatrix} x & y \\ z & w \end{bmatrix} &= \mathbf{v} \\ &= T(\mathbf{u}) \\ &= T\left(a + bx + cx^2 + dx^3\right) \\ &= \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix} \end{aligned}$$

Matrix equality leads us to the system of four equations in the four unknowns, x, y, z, w,

$$\begin{aligned} a + b &= x \\ a - 2c &= y \\ d &= z \\ b - d &= w \end{aligned}$$

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

$$\begin{aligned} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} &= \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \\ &= \begin{bmatrix} x - z - w \\ z + w \\ \frac{1}{2}(x - y - z - w) \\ z \end{bmatrix} \end{aligned}$$

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 [411] that T is surjective. All the same, lets do a concrete demonstration and evaluate T with $\mathbf{u}$,

$$\begin{aligned} T(\mathbf{u}) &= T\left((x - z - w) + (z + w)x + \frac{1}{2}(x - y - z - w)x^2 + zx^3\right) \\ &= \begin{bmatrix} (x - z - w) + (z + w) & (x - z - w) - 2(\frac{1}{2}(x - y - z - w)) \\ z & (z + w) - z \end{bmatrix} \\ &= \begin{bmatrix} x & y \\ z & w \end{bmatrix} \\ &= \mathbf{v} \end{aligned}$$

⊙

Subsection RLT
Range of a Linear Transformation

For a linear transformation $T\colon U \mapsto V$, 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 range 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 $T\colon U \mapsto V$ is a linear transformation. Then the **range** of T is the set

$$\mathcal{R}(T) = \{T(\mathbf{u}) \mid \mathbf{u} \in U\}$$

△

Example RAO
Range, Archetype O
Archetype O [533] is the linear transformation

$$T\colon \mathbb{C}^3 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -x_1 + x_2 - 3x_3 \\ -x_1 + 2x_2 - 4x_3 \\ x_1 + x_2 + x_3 \\ 2x_1 + 3x_2 + x_3 \\ x_1 + 2x_3 \end{bmatrix}$$

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 $\mathbf{u} \in \mathbb{C}^3$,

$$\begin{aligned}
\mathbf{v} &= T(\mathbf{u}) \\
&= \begin{bmatrix} -u_1 + u_2 - 3u_3 \\ -u_1 + 2u_2 - 4u_3 \\ u_1 + u_2 + u_3 \\ 2u_1 + 3u_2 + u_3 \\ u_1 + 2u_3 \end{bmatrix} \\
&= \begin{bmatrix} -u_1 \\ -u_1 \\ u_1 \\ 2u_1 \\ u_1 \end{bmatrix} + \begin{bmatrix} u_2 \\ 2u_2 \\ u_2 \\ 3u_2 \\ 0 \end{bmatrix} + \begin{bmatrix} -3u_3 \\ -4u_3 \\ u_3 \\ u_3 \\ 2u_3 \end{bmatrix} \\
&= u_1 \begin{bmatrix} -1 \\ -1 \\ 1 \\ 2 \\ 1 \end{bmatrix} + u_2 \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 0 \end{bmatrix} + u_3 \begin{bmatrix} -3 \\ -4 \\ 1 \\ 1 \\ 2 \end{bmatrix}
\end{aligned}$$

This says that every output of $T(\mathbf{v})$ can be written as a linear combination of the three vectors

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

using the scalars u_1, u_2, u_3. Furthermore, since $\mathbf{u}$ can be any element of $\mathbb{C}^3$, every such linear combination is an output. This means that

$$\mathcal{R}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix}-1\\-1\\1\\2\\1\end{bmatrix}, \begin{bmatrix}1\\2\\1\\3\\0\end{bmatrix}, \begin{bmatrix}-3\\-4\\1\\1\\2\end{bmatrix}\right\}\right)$$

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 RM [167] and Section RSM [183]. We will place the vectors into a matrix as rows, row-reduce, toss out zero rows and appeal to Theorem BRS [186], so we can describe the range of T with a basis,

$$\mathcal{R}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\-3\\-7\\-2\end{bmatrix}, \begin{bmatrix}0\\1\\2\\5\\1\end{bmatrix}\right\}\right)$$ ⊚

We know that the span of a set of vectors is always a subspace (Theorem SSS [260]), so the range computed in Example RAO [416] 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 $T\colon U \mapsto V$ 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 [255]. First $T(\mathbf{0}_U) = \mathbf{0}_V$ by Theorem LTTZZ [383], 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 implies that $\mathbf{w} + \mathbf{z} \in U$. Then

$$\begin{aligned} T(\mathbf{w}+\mathbf{z}) &= T(\mathbf{w}) + T(\mathbf{z}) && \text{Definition LT [379]} \\ &= \mathbf{x} + \mathbf{y} && \text{Definition of } \mathbf{w} \text{ and } \mathbf{z} \end{aligned}$$

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

$$\begin{aligned} T(\alpha\mathbf{w}) &= \alpha T(\mathbf{w}) && \text{Definition LT [379]} \\ &= \alpha\mathbf{x} && \text{Definition of } \mathbf{w} \end{aligned}$$ ■

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 [255] 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 [531] is the linear transformation

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^3, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 + 3x_3 - 4x_4 + 5x_5 \\ x_1 - 2x_2 + 3x_3 - 9x_4 + 3x_5 \\ 3x_1 + 4x_3 - 6x_4 + 5x_5 \end{bmatrix}$$

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 $\mathbf{u} \in \mathbb{C}^5$,

$$\begin{aligned} \mathbf{v} &= T(\mathbf{u}) \\ &= \begin{bmatrix} 2u_1 + u_2 + 3u_3 - 4u_4 + 5u_5 \\ u_1 - 2u_2 + 3u_3 - 9u_4 + 3u_5 \\ 3u_1 + 4u_3 - 6u_4 + 5u_5 \end{bmatrix} \\ &= \begin{bmatrix} 2u_1 \\ u_1 \\ 3u_1 \end{bmatrix} + \begin{bmatrix} u_2 \\ -2u_2 \\ 4u_2 \end{bmatrix} + \begin{bmatrix} 3u_3 \\ 3u_3 \\ 4u_3 \end{bmatrix} + \begin{bmatrix} -4u_4 \\ -9u_4 \\ -6u_4 \end{bmatrix} + \begin{bmatrix} 5u_5 \\ 3u_5 \\ 5u_5 \end{bmatrix} \\ &= u_1\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} + u_2\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} + u_3\begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix} + u_4\begin{bmatrix} -4 \\ -9 \\ -6 \end{bmatrix} + u_5\begin{bmatrix} 5 \\ 3 \\ 5 \end{bmatrix} \end{aligned}$$

This says that every output of T $(\mathbf{v})$ can be written as a linear combination of the five vectors

$$\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \qquad \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} \qquad \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix} \qquad \begin{bmatrix} -4 \\ -9 \\ -6 \end{bmatrix} \qquad \begin{bmatrix} 5 \\ 3 \\ 5 \end{bmatrix}$$

using the scalars u_1, u_2, u_3, u_4, u_5. Furthermore, since $\mathbf{u}$ can be any element of $\mathbb{C}^5$, every such linear combination is an output. This means that

$$\mathcal{R}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\1\\3\end{bmatrix},\,\begin{bmatrix}1\\-2\\4\end{bmatrix},\,\begin{bmatrix}3\\3\\4\end{bmatrix},\,\begin{bmatrix}-4\\-9\\-6\end{bmatrix},\,\begin{bmatrix}5\\3\\5\end{bmatrix}\right\}\right)$$

The five vectors in this spanning set for $\mathcal{R}(T)$ form a linearly dependent set (Theorem MVSLD [133]). So we can find a more economical presentation by any of the various methods from Section RM [167] and Section RSM [183]. We will place the vectors into a matrix as rows, row-reduce, toss out zero rows and appeal to Theorem BRS [186], so we can describe the range of T with a (nice) basis,

$$\mathcal{R}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$ ⊚

In contrast to injective linear transformations having small (trivial) null spaces (Theorem NSILT [406]), surjective linear transformations have large ranges, as indicated in the next theorem.

Theorem RSLT
Range of a Surjective Linear Transformation
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then T is surjective if and only if the range of T equals the codomain, $\mathcal{R}(T) = V$. □

Proof (⇒) By Definition RLT [416], 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 [411]), 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 [411]. In that example, we showed that Archetype Q [538] 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 [411]. 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

$$\mathcal{R}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\\-1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\\-1\end{bmatrix}, \begin{bmatrix}0\\0\\0\\1\\2\end{bmatrix}\right\}\right)$$

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

$$\mathbf{v}^* = (-1)\begin{bmatrix}1\\0\\0\\0\\1\end{bmatrix} + (2)\begin{bmatrix}0\\1\\0\\0\\-1\end{bmatrix} + (3)\begin{bmatrix}0\\0\\1\\0\\-1\end{bmatrix} + (-1)\begin{bmatrix}0\\0\\0\\1\\2\end{bmatrix} = \begin{bmatrix}-1\\2\\3\\-1\\-8\end{bmatrix}$$

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 ($\mathbf{v}$ in Example NSAQ [411]) outside of the range. Any attempt to find an input for T that will produce $\mathbf{v}$ 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 [419]. 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 [416] the range of Archetype O [533] was determined to be

$$\mathcal{R}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\-3\\-7\\-2\end{bmatrix}, \begin{bmatrix}0\\1\\2\\5\\1\end{bmatrix}\right\}\right)$$

a subspace of dimension 2 in $\mathbb{C}^5$. Since $\mathcal{R}(T) \neq \mathbb{C}^5$, Theorem RSLT [419] says T is not surjective. ⊚

Example SAN
Surjective, Archetype N

The range of Archetype N [531] was computed in Example FRAN [418] to be

$$\mathcal{R}(T) = \left\{\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right\}$$

Since the basis for this subspace is the set of standard unit vectors for $\mathbb{C}^3$ (Theorem SUVB [280]), we have $\mathcal{R}(T) = \mathbb{C}^3$ and by Theorem RSLT [419], T is surjective. ⊚

Just as injective linear transformations are allied with linear independence (Theorem ILTLI [408], Theorem ILTB [408]), surjective linear transformations are allied with spanning sets.

Theorem SLTS
Surjective Linear Transformations and Spans
Suppose that $T\colon U \mapsto V$ is a surjective linear transformation and $S = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_t\}$ spans U. Then $R = \{T(\mathbf{u}_1),\, T(\mathbf{u}_2),\, T(\mathbf{u}_3),\, \ldots,\, T(\mathbf{u}_t)\}$ spans $\mathcal{R}(T)$. □

Proof We need to establish that every element of $\mathcal{R}(T)$ can be written as a linear combination of the vectors in R. 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 SLT [411]).

Because S spans V there are scalars, $a_1,\, a_2,\, a_3,\, \ldots,\, a_t$, such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t$$

Then

$$\begin{aligned}
\mathbf{v} &= T(\mathbf{u}) && T \text{ surjective}\\
&= T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_t\mathbf{u}_t) && S \text{ spans } V\\
&= a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) + a_3T(\mathbf{u}_3) + \ldots + a_tT(\mathbf{u}_t) && \text{Theorem LTLC [389]}
\end{aligned}$$

which establishes that R spans the range of T, $\mathcal{R}(T)$. ■

TODO: Great way to span range.

Elements of the range are precisely those elements of the codomain with non-empty preimages.

Theorem RPI
Range and Pre-Image
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then

$$\mathbf{v} \in \mathcal{R}(T) \text{ if and only if } T^{-1}(\mathbf{v}) \neq \emptyset$$ □

Proof (⇒) 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.

(⇐) Suppose the preimage of $\mathbf{v}$ 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 $T\colon U \mapsto V$ 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 (⇒) Assume T is surjective. Since B is a basis, we know B is a spanning set of U (Definition B [279]). Then Theorem ILTLI [408] says that C spans $\mathcal{R}(T)$. But the hypothesis that T is surjective means $V = \mathcal{R}(T)$ (Theorem RSLT [419]), so C spans V.

(⇐) 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 [411]). 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

$$\mathbf{v} = b_1 T(\mathbf{u}_1) + b_2 T(\mathbf{u}_2) + b_3 T(\mathbf{u}_3) + \cdots + b_m T(\mathbf{u}_m)$$

Now define the vector $\mathbf{u} \in U$ by

$$\mathbf{u} = b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 + \cdots + b_m\mathbf{u}_m$$

Then

$$\begin{aligned} T(\mathbf{u}) &= T(b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 + \cdots + b_m\mathbf{u}_m) \\ &= b_1 T(\mathbf{u}_1) + b_2 T(\mathbf{u}_2) + b_3 T(\mathbf{u}_3) + \cdots + b_m T(\mathbf{u}_m) && \text{Theorem LTLC [389]} \\ &= \mathbf{v} \end{aligned}$$

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 [411], T is surjective. ■

Subsection SLTD
Surjective Linear Transformations and Dimension

Theorem SLTD
Surjective Linear Transformations and Dimension
Suppose that $T\colon U \mapsto V$ is a surjective linear transformation. Then $\dim(U) \geq \dim(V)$. □

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 [421], 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 [310], so our assumption is false and $\dim(U) \geq \dim(V)$. ■

Example NSDAT
Not surjective by dimension, Archetype T
The linear transformation in Archetype T [543] is

$$T\colon P_4 \mapsto P_5, \quad T(p(x)) = (x-2)p(x)$$

Since $\dim(P_4) = 5 < 6 = \dim(P_5)$, T cannot be surjective for then it would violate Theorem SLTD [422]. ⊚

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 [533] and Archetype P [536] 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 [394] we saw how to combine linear transformations to build new linear transformations, specifically, how to build the composition of two linear transformations (Definition LTC [397]). 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 $T\colon U \mapsto V$ and $S\colon V \mapsto W$ are surjective linear transformations. Then $(S \circ T)\colon U \mapsto W$ is a surjective linear transformation. □

Proof That the composition is a linear transformation was established in Theorem CLTLT [397], so we need only establish that the composition is surjective. Applying Definition SLT [411], 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 injective guarantees a vector $\mathbf{u} \in U$ such that $T(\mathbf{u}) = \mathbf{v}$. Now,

$$\begin{aligned}
(S \circ T)(\mathbf{u}) &= S(T(\mathbf{u})) && \text{Definition LTC [397]}\\
&= S(\mathbf{v}) && T \text{ surjective}\\
&= \mathbf{w} && S \text{ surjective}
\end{aligned}$$

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 [411], $S \circ T$ is surjective. ■

Subsection READ
Reading Questions

1. Suppose $T\colon \mathbb{C}^5 \mapsto \mathbb{C}^8$ is a linear transformation. Why can't T be surjective?
2. What is the relationship between a surjective linear transformation and its range?
3. Compare and contrast injective and surjective linear transformations.

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

$$I_W\colon W \mapsto W, \qquad I_W(\mathbf{w}) = \mathbf{w}$$ △

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 null space and range to see that it is both injective and surjective. All of these facts should be straightforward to verify. With this in hand we can make our main definition.

Definition IVLT
Invertible Linear Transformations
Suppose that $T\colon U \mapsto V$ is a linear transformation. If there is a function $S\colon V \mapsto U$ 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 analgous to squaring a 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 [425].

Example AIVLT
An invertible linear transformation

Archetype V [544] is the linear transformation

$$T\colon P_3 \mapsto M_{22},\quad T\left(a+bx+cx^2+dx^3\right)=\begin{bmatrix}a+b & a-2c\\ d & b-d\end{bmatrix}$$

Define the function $S\colon M_{22}\mapsto P_3$ 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

$$\begin{aligned}(T\circ S)\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right)&=T\left(S\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right)\right)\\
&=T\left((a-c-d)+(c+d)x+\frac{1}{2}(a-b-c-d)x^2+cx^3\right)\\
&=\begin{bmatrix}(a-c-d)+(c+d) & (a-c-d)-2(\frac{1}{2}(a-b-c-d))\\ c & (c+d)-c\end{bmatrix}\\
&=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\\
&=I_{M_{22}}\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right)\end{aligned}$$

And

$$\begin{aligned}(S\circ T)\left(a+bx+cx^2+dx^3\right)&=S\left(T\left(a+bx+cx^2+dx^3\right)\right)\\
&=S\left(\begin{bmatrix}a+b & a-2c\\ d & b-d\end{bmatrix}\right)\\
&=((a+b)-d-(b-d))+(d+(b-d))x\\
&\quad+\left(\frac{1}{2}((a+b)-(a-2c)-d-(b-d))\right)x^2+(d)x^3\\
&=a+bx+cx^2+dx^3\\
&=I_{P_3}\left(a+bx+cx^2+dx^3\right)\end{aligned}$$

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 $T\colon \mathbb{C}^3\mapsto M_{22}$ 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 $S\colon M_{22} \mapsto \mathbb{C}^3$.

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

$$\begin{aligned} a - b &= 5 \\ 2a + 2b + c &= 3 \\ 3a + b + c &= 8 \\ -2a - 6b - 2c &= 2 \end{aligned}$$

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

$$T\left(\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 3 & 2 \\ 5 & 2 \end{bmatrix} = B \qquad\qquad T\left(\begin{bmatrix} 0 \\ -3 \\ 8 \end{bmatrix}\right) = \begin{bmatrix} 3 & 2 \\ 5 & 2 \end{bmatrix} = B$$

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 null space of T,

$$\mathcal{N}(T) = \mathcal{S}p\left(\left\{\begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}\right\}\right)$$

Now choose either of the two inputs used above for T and add to it a scalar multiple of the basis vector for the null space 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. ⊚

In Example ANILT [426] 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 inverse putative 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 multiply-defined on each of their inputs. We will formalize these ideas in Theorem ILTIS [429].

But first notice in Definition IVLT [425] 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
Invese of a Linear Transformation is a Linear Transformation
Suppose that $T\colon U\mapsto V$ is an invertible linear transformation. Then the function $T^{-1}\colon V\mapsto U$ is a linear transformation. □

Proof We work through verifying Definition LT [379] for T^{-1}, employing as we go properties of T given by Definition LT [379]. To this end, suppose $\mathbf{x},\,\mathbf{y}\in V$ and $\alpha\in\mathbb{C}$.

$$\begin{aligned}
T^{-1}(\mathbf{x}+\mathbf{y}) &= T^{-1}\left(T\left(T^{-1}(\mathbf{x})\right)+T\left(T^{-1}(\mathbf{y})\right)\right) && T,\ T^{-1}\text{ inverse functions}\\
&= T^{-1}\left(T\left(T^{-1}(\mathbf{x})+T^{-1}(\mathbf{y})\right)\right) && T\text{ a linear transformation}\\
&= T^{-1}(\mathbf{x})+T^{-1}(\mathbf{y}) && T^{-1},\ T\text{ inverse functions}
\end{aligned}$$

Now check the second defining property of a linear transformation for T^{-1},

$$\begin{aligned}
T^{-1}(\alpha\mathbf{x}) &= T^{-1}\left(\alpha T\left(T^{-1}(\mathbf{x})\right)\right) && T,\ T^{-1}\text{ inverse functions}\\
&= T^{-1}\left(T\left(\alpha T^{-1}(\mathbf{x})\right)\right) && T\text{ a linear transformation}\\
&= \alpha T^{-1}(\mathbf{x}) && T^{-1},\ T\text{ inverse functions}
\end{aligned}$$ ■

So T^{-1} fullfills the requirements of Definition LT [379] 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 $T\colon U\mapsto V$ is an invertible linear transformation. Then T^{-1} is an invertible linear transformation and $\left(T^{-1}\right)^{-1}=T$. □

Proof Because T is invertible, Definition IVLT [425] tells us there is a function $T^{-1}\colon V \mapsto U$ such that

$$T^{-1} \circ T = I_U \qquad\qquad T \circ T^{-1} = I_V$$

Additionally, Theorem ILTLT [428] 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 [425], 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. $\left(T^{-1}\right)^{-1} = T$. ■

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.

Theorem ILTIS
Invertible Linear Transformations are Injective and Surjective
Suppose $T\colon U \mapsto V$ is a linear transformation. Then T is invertible if and only if T is injective and surjective. □

Proof (⇒) Since T is presumed invertible, we can employ its inverse, T^{-1} (Definition IVLT [425]). To see that T is injective, suppose $\mathbf{x}, \mathbf{y} \in U$

$$\begin{aligned}
T(\mathbf{x}) &= T(\mathbf{y}) && \text{Definition ILT [399]}\\
T^{-1}(T(\mathbf{x})) &= T^{-1}(T(\mathbf{y})) && \text{Apply } T^{-1}\\
\left(T^{-1} \circ T\right)(\mathbf{x}) &= \left(T^{-1} \circ T\right)(\mathbf{y}) && \text{Definition LTC [397]}\\
I_U(\mathbf{x}) &= I_U(\mathbf{y}) && T^{-1}, T \text{ inverse functions}\\
\mathbf{x} &= \mathbf{y} && \text{Definition IDLT [425]}
\end{aligned}$$

So by Definition ILT [399] T is injective. To check that T is surjective, suppose $\mathbf{v} \in V$. Employ T^{-1} by defining $\mathbf{u} = T^{-1}(\mathbf{v})$. Then

$$\begin{aligned}
T(\mathbf{u}) &= T\left(T^{-1}(\mathbf{v})\right) && \text{Substitution for } \mathbf{u}\\
&= \left(T \circ T^{-1}\right)(\mathbf{v}) && \text{Definition LTC [397]}\\
&= I_V(\mathbf{v}) && T, T^{-1} \text{ inverse functions}\\
&= \mathbf{v} && \text{Definition IDLT [425]}
\end{aligned}$$

So there is an input to T, $\mathbf{u}$, that produces the chosen output, $\mathbf{v}$, and hence T is surjective by Definition SLT [411].

(⇐) Now assume that T is both injective and surjective. We will build a function $S\colon V \mapsto U$ that will establish that T is invertible. To this end, choose any $\mathbf{v} \in V$.

Since T is surjective, Theorem RSLT [419] says $\mathcal{R}(T) = V$, so we have $\mathbf{v} \in \mathcal{R}(T)$. Theorem RPI [421] 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 NSPI [406] then says that

$$T^{-1}(\mathbf{v}) = \{\mathbf{u} + \mathbf{z} \mid \mathbf{z} \in \mathcal{N}(T)\}$$

However, because T is injective, by Theorem NSILT [406] the null space is trivial, $\mathcal{N}(T) = \{\mathbf{0}\}$. 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 $\mathbf{u}$ was arbitary 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 $\mathbf{v}$ was arbitary we have function equality, $T \circ S = I_V$.

We will make frequent use of this characterization of invertible linear transformations. 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 $T\colon U \mapsto V$ and $S\colon V \mapsto W$ are invertible linear transformations. Then the composition, $(S \circ T) : U \mapsto W$ 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 [397]. Since S and T are both invertible, Theorem ILTIS [429] says that S and T are both injective and surjective. Then Theorem CILTI [410] says $S \circ T$ is injective, and Theorem CSLTS [423] says $S \circ T$ is surjective. Now apply the "other half" of Theorem ILTIS [429] 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 $T\colon U \mapsto V$ and $S\colon V \mapsto W$ are invertible linear transformations. Then $S \circ T$ is invertible and $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$. □

Proof Compute, for all $\mathbf{w} \in W$

$$\begin{aligned}
\left((S \circ T) \circ \left(T^{-1} \circ S^{-1}\right)\right)(\mathbf{w}) &= S\left(T\left(T^{-1}\left(S^{-1}(\mathbf{w})\right)\right)\right) \\
&= S\left(I_V\left(S^{-1}(\mathbf{w})\right)\right) && \text{Definition IVLT [425]} \\
&= S\left(S^{-1}(\mathbf{w})\right) && \text{Definition IDLT [425]} \\
&= \mathbf{w} && \text{Definition IVLT [425]} \\
&= I_W(\mathbf{w}) && \text{Definition IDLT [425]}
\end{aligned}$$

so $(S \circ T) \circ (T^{-1} \circ S^{-1}) = I_W$ and also

$$\begin{aligned}
\left(\left(T^{-1} \circ S^{-1}\right) \circ (S \circ T)\right)(\mathbf{u}) &= T^{-1}\left(S^{-1}\left(S\left(T\left(\mathbf{u}\right)\right)\right)\right) \\
&= T^{-1}\left(I_V\left(T\left(\mathbf{u}\right)\right)\right) && \text{Definition IVLT [425]} \\
&= T^{-1}\left(T\left(\mathbf{u}\right)\right) && \text{Definition IDLT [425]} \\
&= \mathbf{u} && \text{Definition IVLT [425]} \\
&= I_U(\mathbf{u}) && \text{Definition IDLT [425]}
\end{aligned}$$

so $(T^{-1} \circ S^{-1}) \circ (S \circ T) = I_U$. By Definition IVLT [425], $S \circ T$ 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 [430] and also establishes the invertibility of $S \circ T$. But somehow, the proof of Theorem CIVLT [430] is nicer way to get this property.

Does this Theorem ICLT [431] 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 [239]) as a set of objects ("vectors") endowed with a definition of vector addition (+) and a definition of scalar multiplication (juxtaposition). Many of our definitions about vector spaces involve linear combinations (Definition LC [259]), such as the span of a set (Definition SS [260]) and linear independence (Definition LI [271]). Other definitions are built up from these ideas, such as bases (Definition B [279]) and dimension (Definition D [293]). 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 [379]). 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 axioms of Definition VS [239]. 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, $T\colon U \mapsto V$. In this case, we write $U \cong V$, and the linear transformation T is known as an **isomorphism** between U and V. △

A few comments on this definition. First, be careful with your language (Technique L [26]). Two vector spaces are isomrphic, 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 [367] for a reminder about equivalence relations).

Example IVSAV
Isomorphic vector spaces, Archetype V
Archetype V [544] is a linear transformation from P_3 to M_{22},

$$T\colon P_3 \mapsto M_{22}, \quad T\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

Since it is injective and surjective, Theorem ILTIS [429] tells us that it is an invertible linear transformation. By Definition IVS [432] 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 axioms there (Example VSM [241]),

$$\begin{aligned}
&T\left(5(2+3x-4x^2+5x^3)+(-3)(3-5x+3x^2+x^3)\right) \\
&= 5T\left(2+3x-4x^2+5x^3\right)+(-3)T\left(3-5x+3x^2+x^3\right) && \text{Theorem LTLC [389]} \\
&= 5\begin{bmatrix}5 & 10\\ 5 & -2\end{bmatrix}+(-3)\begin{bmatrix}-2 & -3\\ 1 & -6\end{bmatrix} && \text{Definition of } T \\
&= \begin{bmatrix}31 & 59\\ 22 & 8\end{bmatrix} && \text{Operations in } M_{22}
\end{aligned}$$

Now we will translate our answer back to P_3 by appling T^{-1}, which we found in Example AIVLT [426],

$$T^{-1}\colon M_{22}\mapsto P_3,\quad T^{-1}\left(\begin{bmatrix}a & b\\ c & d\end{bmatrix}\right)=(a-c-d)+(c+d)x+\frac{1}{2}(a-b-c-d)x^2+cx^3$$

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 [242]).⊚

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 $T\colon U\mapsto V$ (Definition IVS [432]). T is injective by Theorem ILTIS [429] and so by Theorem ILTD [409], $\dim(U)\leq\dim(V)$. Similarly, T is surjective by Theorem ILTIS [429] and so by Theorem SLTD [422], $\dim(U)\geq\dim(V)$. The net effect of these two inequalities is that $\dim(U)=\dim(V)$. ■

The contrapositive of Theorem IVSED [433] 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 [441] we will be able to establish that the converse of Theorem IVSED [433] 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 [302]) the rank and nullity of a linear transformation are related. Here are the definitions and theorems, see the Archetypes (Chapter A [469]) for loads of examples.

Definition ROLT
Rank Of a Linear Transformation
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then the **rank** of T, $r\left(T\right)$, is the dimension of the range of T,

$$r\left(T\right) = \dim\left(\mathcal{R}(T)\right)$$ △

Definition NOLT
Nullity Of a Linear Transformation
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then the **nullity** of T, $n\left(T\right)$, is the dimension of the null space of T,

$$n\left(T\right) = \dim\left(\mathcal{N}(T)\right)$$ △

Here are two quick theorems.

Theorem ROSLT
Rank Of a Surjective Linear Transformation
Suppose that $T\colon U \mapsto V$ is a linear transformation. Then the rank of T is the dimension of V, $r\left(T\right) = \dim\left(V\right)$, if and only if T is surjective. □

Proof By Theorem RSLT [419], T is surjective if and only if $\mathcal{R}(T) = V$. Applying Definition ROLT [434], $\mathcal{R}(T) = V$ if and only if $r\left(T\right) = \dim\left(\mathcal{R}(T)\right) = \dim\left(V\right)$. ■

Theorem NOILT
Nullity Of an Injective Linear Transformation
Suppose that $T\colon U \mapsto V$ is an injective linear transformation. Then the nullity of T is zero, $n\left(T\right) = 0$, if and only if T is injective. □

Proof By Theorem NSILT [406], T is injective if and only if $\mathcal{N}(T) = \{\mathbf{0}\}$. Applying Definition NOLT [434], $\mathcal{N}(T) = \{\mathbf{0}\}$ if and only if $n\left(T\right) = 0$. ■

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 $T\colon U \mapsto V$ is a linear transformation. Then

$$r\left(T\right) + n\left(T\right) = \dim\left(U\right)$$ □

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 null space of T, $\mathcal{N}(T)$. 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 pre-image by Theorem RPI [421]. 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

$$B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_s, \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_r\}$$

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 $b_1, b_2, b_3, \ldots, b_r$

$$\mathbf{0} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_s\mathbf{u}_s + b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + b_3\mathbf{w}_3 + \cdots + b_r\mathbf{w}_r$$

Then

$$\begin{aligned}
\mathbf{0} &= T(\mathbf{0}) && \text{Theorem LTTZZ [383]}\\
&= T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_s\mathbf{u}_s\\
&\qquad + b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + b_3\mathbf{w}_3 + \cdots + b_r\mathbf{w}_r) && \text{Substitution}\\
&= a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) + a_3T(\mathbf{u}_3) + \cdots + a_sT(\mathbf{u}_s)\\
&\quad + b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \cdots + b_rT(\mathbf{w}_r) && \text{Theorem LTLC [389]}\\
&= a_1\mathbf{0} + a_2\mathbf{0} + a_3\mathbf{0} + \cdots + a_s\mathbf{0}\\
&\quad + b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \cdots + b_rT(\mathbf{w}_r) && \mathbf{u}_i \in \mathcal{N}(T)\\
&= \mathbf{0} + \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0}\\
&\quad + b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \cdots + b_rT(\mathbf{w}_r) && \text{Theorem ZVSM [247]}\\
&= b_1T(\mathbf{w}_1) + b_2T(\mathbf{w}_2) + b_3T(\mathbf{w}_3) + \cdots + b_rT(\mathbf{w}_r) && \text{Property zero vector}\\
&= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + \cdots + b_r\mathbf{v}_r && \mathbf{w}_i \in T^{-1}(\mathbf{v}_i)
\end{aligned}$$

This is a relation of linear dependence on R (Definition RLD [271]), and since R is a linearly independent set (Definition LI [271]), we see that $b_1 = b_2 = b_3 = \ldots = b_r = 0$. Then the original relation of linear dependence on B becomes

$$\begin{aligned}
\mathbf{0} &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_s\mathbf{u}_s + 0\mathbf{w}_1 + 0\mathbf{w}_2 + \ldots + 0\mathbf{w}_r\\
&= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_s\mathbf{u}_s + \mathbf{0} + \mathbf{0} + \ldots + \mathbf{0} && \text{Theorem ZSSM [247]}\\
&= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_s\mathbf{u}_s && \text{Property zero vector}
\end{aligned}$$

But this is again a relation of linear independence (Definition RLD [271]), now on the set S. Since S is linearly independent (Definition LI [271]), 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 [271].

To now establish that B spans U, choose an arbitrary vector $\mathbf{u} \in U$. Then $T(\mathbf{u}) \in \mathcal{R}(T)$, so there are scalars $c_1, c_2, c_3, \ldots, c_r$ such that

$$T(\mathbf{u}) = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_r\mathbf{v}_r$$

Use the scalars $c_1, c_2, c_3, \ldots, c_r$ to define a vector $\mathbf{y} \in U$,

$$\mathbf{y} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + \cdots + c_r\mathbf{w}_r$$

Then

$$\begin{aligned}
T(\mathbf{u}-\mathbf{y}) &= T(\mathbf{u}) - T(\mathbf{y}) && \text{Theorem LTLC [389]}\\
&= T(\mathbf{u}) - T(c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + \cdots + c_r\mathbf{w}_r) && \text{Substitution}\\
&= T(\mathbf{u}) - (c_1T(\mathbf{w}_1) + c_2T(\mathbf{w}_2) + \cdots + c_rT(\mathbf{w}_r)) && \text{Theorem LTLC [389]}\\
&= T(\mathbf{u}) - (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_r\mathbf{v}_r) && \mathbf{w}_i \in T^{-1}(\mathbf{v}_i)\\
&= T(\mathbf{u}) - T(\mathbf{u}) && \text{Substitution}\\
&= \mathbf{0} && \text{Additive inverses}
\end{aligned}$$

So the vector $\mathbf{u}-\mathbf{y}$ is sent to the zero vector by T and hence is an element of the null space of T. As such it can be written as a linear combination of the basis vectors for $\mathcal{N}(T)$, the elements of the set S. So there are scalars $d_1, d_2, d_3, \ldots, d_s$ such that

$$\mathbf{u}-\mathbf{y} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + d_3\mathbf{u}_3 + \cdots + d_s\mathbf{u}_s$$

Then

$$\begin{aligned}
\mathbf{u} &= (\mathbf{u}-\mathbf{y}) + \mathbf{y}\\
&= d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + d_3\mathbf{u}_3 + \cdots + d_s\mathbf{u}_s + c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + \cdots + c_r\mathbf{w}_r
\end{aligned}$$

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 [260]).

So B is a basis (Definition B [279]) of U with $s+r$ vectors, and thus

$$\dim(U) = s + r = n(T) + r(T)$$

as desired. ■

Theorem RPNC [302] 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 [435] when we consider the linear transformation $T\colon \mathbb{C}^n \mapsto \mathbb{C}^m$ defined with the $m \times n$ matrix A by $T(\mathbf{x}) = A\mathbf{x}$. The range and null space of T are identical to the range and null space of the matrix A (can you prove this?), 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 determing basic properties of linear transformations. For example, suppose that $T\colon \mathbb{C}^6 \mapsto \mathbb{C}^6$ is a linear transformation and you are able to quickly establish that the null space is trivial. Then $n(T) = 0$. First this means that T is injective by Theorem NOILT [434]. Also, Theorem RPNDD [435] becomes

$$6 = \dim\left(\mathbb{C}^6\right) = r(T) + n(T) = r(T) + 0 = r(T)$$

So the rank of T is equal to the rank of the codomain, and by Theorem ROSLT [434] we know T is surjective. Finally, we know T is invertible by Theorem ILTIS [429]. So from the determination that the null space 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 [435] can be used to provide alternative proofs for Theorem ILTD [409], Theorem SLTD [422] and Theorem IVSED [433]. 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 the dimensions of the domain and codomain, and possibly with just the nullity or rank. The table preceding all of the archetypes 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. Its 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 [3], 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 [441].

Archetype D [487] and Archetype E [491] 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 [197]) and define the linear transformation,

$$T\colon \mathbb{C}^4 \mapsto \mathbb{C}^3, \quad T(\mathbf{x}) = D\mathbf{x} = x_1 \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 7 \\ -5 \\ 4 \end{bmatrix} + x_4 \begin{bmatrix} -7 \\ -6 \\ -5 \end{bmatrix}$$

Theorem MBLT [385] tells us that T is indeed a linear transformation. Archetype D [487] 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 $T^{-1}\left(\mathbf{b}\right)$. 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 NSPI [406] tells us that the complete solution set of the linear system is the preimage of $\mathbf{b}$,

$$\mathbf{w} + \mathcal{N}(T) = \{\mathbf{w} + \mathbf{z} \mid \mathbf{z} \in \mathcal{N}(T)\}$$

The null space of the linear transformation T is exactly the null space of the matrix D (Theorem XX [**??**]), so this approach to the solution set should be reminiscent of Theorem PSPHS [206]. The null space 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 [491] is identical to Archetype D [487] 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 $\mathbf{d}$, $T^{-1}\left(\mathbf{d}\right)$. However, the vector $\mathbf{d}$ is not in the range of the linear transformation (nor is it in the range of the matrix, since these two ranges are equal sets by Theorem XX [**??**]). 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 [59] 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

$$\begin{aligned} n\left(T\right) &= \dim\left(\mathbb{C}^4\right) - r\left(T\right) && \text{Theorem RPNDD [435]} \\ &= 4 - \dim\left(\mathcal{R}(T)\right) && \text{Definition ROLT [434]} \\ &\geq 4 - 3 && \mathcal{R}(T) \text{ subspace of } \mathbb{C}^3 \\ &= 1 \end{aligned}$$

So the null space 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 NSPI [406] tells us that the pre-images are built from the null space, and with a non-trivial null space, 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 $S\left(\mathbf{x}\right) = C\mathbf{x}$. If S has a trivial null space, 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 with 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 [435] 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 [429] 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 SNSCM [229]), we can use the inverse of the linear transformation to construct the unique element of any pre-image (proof of Theorem ILTIS [429]).

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 [441].

Subsection READ
Reading Questions

1. What conditions allow us to easily determine if a linear tranformation 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?

R: Representations

Section VR
Vector Representations

Definition VR
Vector Representation
Suppose that V is a vector space with a basis $B = \{\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\ldots,\,\mathbf{v}_n\}$. Define a function $\rho_B\colon V \mapsto \mathbb{C}^n$ as follows. For $\mathbf{w} \in V$, find some scalars $a_1,\,a_2,\,a_3,\,\ldots,\,a_n$ so that

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_n\mathbf{v}_n$$

then

$$\rho_B\left(\mathbf{w}\right) = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} \qquad \triangle$$

We need to show that ρ_B is really a function (since "find some scalars" sounds like it could be accomplished in many ways, or perhaps not at all) and right now we want to establish that ρ_B is a linear transformation. We will wrap up both objectives in one theorem, even though the first part is working backwards to make sure that ρ_B is well-defined.

Theorem VRLT
Vector Representation is a Linear Transformation
The function ρ_B (Definition VR [441]) is a linear transformation. □

Proof The definition of ρ_B (Definition VR [441]) appears to allow considerable latitude in selecting the scalars $a_1,\,a_2,\,a_3,\,\ldots,\,a_n$. However, since B is a basis for V, Theorem VRRB [287] says this can be done, and done *uniquely*. So despite appearances, ρ_B is indeed a function.

Suppose that $\mathbf{x}$ and $\mathbf{y}$ are two vectors in V and $\alpha \in \mathbb{C}$. Then the vector space axioms (Definition VS [239]) assure us that the vectors $\mathbf{x} + \mathbf{y}$ and $\alpha\mathbf{x}$ are also vectors in V.

Theorem VRRB [287] then provides the follow sets of scalars for the four vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}+\mathbf{y}$ and $\alpha\mathbf{x}$, and tells us that each set of scalars is the only way to express the given vector as a linear combination of the basis vectors in B.

$$\begin{aligned}
\mathbf{x} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_n\mathbf{v}_n \\
\mathbf{y} &= b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + \cdots + b_n\mathbf{v}_n \\
\mathbf{x}+\mathbf{y} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n \\
\alpha\mathbf{x} &= d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + d_3\mathbf{v}_3 + \cdots + d_n\mathbf{v}_n
\end{aligned}$$

Then

$$\begin{aligned}
&c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n \\
&\quad= \mathbf{x}+\mathbf{y} \\
&\quad= (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_n\mathbf{v}_n) \\
&\qquad+ (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 + \cdots + b_n\mathbf{v}_n) \\
&\quad= a_1\mathbf{v}_1 + b_1\mathbf{v}_1 + a_2\mathbf{v}_2 + b_2\mathbf{v}_2 \\
&\qquad+ a_3\mathbf{v}_3 + b_3\mathbf{v}_3 + \cdots + a_n\mathbf{v}_n + b_n\mathbf{v}_n && \text{Commutativity in } V \\
&\quad= (a_1+b_1)\mathbf{v}_1 + (a_2+b_2)\mathbf{v}_2 \\
&\qquad+ (a_3+b_3)\mathbf{v}_3 + \cdots + (a_n+b_n)\mathbf{v}_n && \text{Distributivity in } V
\end{aligned}$$

By the uniqueness of the expression of $\mathbf{x}+\mathbf{y}$ as a linear combination of the vectors in B, we conclude that $c_i = a_i + b_i$, $1 \leq i \leq n$. So employing our definition of vector addition in $\mathbb{C}^m$ (Definition CVA [91]), we have

$$\rho_B(\mathbf{x}+\mathbf{y}) = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} a_1+b_1 \\ a_2+b_2 \\ a_3+b_3 \\ \vdots \\ a_n+b_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} = \rho_B(\mathbf{x}) + \rho_B(\mathbf{y})$$

so we have the first necessary property for ρ_B to be a linear transformation (Definition LT [379]) . Similarly,

$$\begin{aligned}
&d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + d_3\mathbf{v}_3 + \cdots + d_n\mathbf{v}_n \\
&\quad= \alpha\mathbf{x} \\
&\quad= \alpha(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_n\mathbf{v}_n) \\
&\quad= \alpha(a_1\mathbf{v}_1) + \alpha(a_2\mathbf{v}_2) + \alpha(a_3\mathbf{v}_3) + \cdots + \alpha(a_n\mathbf{v}_n) && \text{Distributivity in } V \\
&\quad= (\alpha a_1)\mathbf{v}_1 + (\alpha a_2)\mathbf{v}_2 + (\alpha a_3)\mathbf{v}_3 + \cdots + (\alpha a_n)\mathbf{v}_n && \text{Associativity in } V
\end{aligned}$$

By the uniqueness of the expression of $\alpha\mathbf{x}$ as a linear combination of the vectors in B, we conclude that $d_i = \alpha a_i$, $1 \leq i \leq n$. So employing our definition of scalar multiplication in $\mathbb{C}^m$ (Definition CVSM [92]), we have

$$\rho_B(\alpha\mathbf{x}) = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_n \end{bmatrix} = \begin{bmatrix} \alpha a_1 \\ \alpha a_2 \\ \alpha a_3 \\ \vdots \\ \alpha a_n \end{bmatrix} = \alpha \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} = \alpha\rho_B(\mathbf{x})$$

and so, with the second property of a linear transformation (Definition LT [379]) confirmed we can conclude that ρ_B is a linear transformation. ■

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 coordinate 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 column of a matrix, checking that the matrix is nonsingular and applying Theorem CNSMB [285]. 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 [101] 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 $\mathbf{y}$. With a nonsingular coefficient matrix, the solution is unique, but this is no surprise as this is the content of Theorem VRRB [287]. This unique solution is

$$a_1 = 2 \qquad a_2 = -1 \qquad a_3 = -3 \qquad a_4 = 4$$

Then by Definition VR [441], we have

$$\rho_B(\mathbf{y}) = \begin{bmatrix} 2 \\ -1 \\ -3 \\ 4 \end{bmatrix}$$

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 $\mathbf{y}$ 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 [441] gives

$$\rho_C(\mathbf{y}) = \begin{bmatrix} -28 \\ -8 \\ 11 \\ 0 \end{bmatrix}$$

We often perform representations relative to standard bases, but for vectors in $\mathbb{C}^m$ its a little silly. Let's find the vector representation of $\mathbf{y}$ relative to the standard basis (Theorem SUVB [280]),

$$D = \{\mathbf{e}_1,\, \mathbf{e}_2,\, \mathbf{e}_3,\, \mathbf{e}_4\}$$

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 [441],

$$\rho_C(\mathbf{y}) = \begin{bmatrix} 6 \\ 14 \\ 6 \\ 7 \end{bmatrix}$$

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 [441],

$$\rho_E(\mathbf{y}) = \begin{bmatrix} 6 \\ 7 \\ 14 \\ 6 \end{bmatrix}$$

So for every basis we could find for $\mathbb{C}^4$ we could construct a representation of $\mathbf{y}$. ⊚

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 [242]). 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 [441]

$$\rho_B(\mathbf{u}) = \begin{bmatrix} 15 \\ 10 \\ -6 \end{bmatrix}$$

Another nice basis for P_2 is

$$B = \left\{1,\, 1+x,\, 1+x+x^2\right\}$$

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,

$$\begin{aligned} 15 &= a_1 + a_2 + a_3 \\ 10 &= a_2 + a_3 \\ -6 &= a_3 \end{aligned}$$

The coefficient matrix of this sytem is nonsingular, leading to a unique solution (no surprise there, see Theorem VRRB [287]),

$$a_1 = 5 \qquad a_2 = 16 \qquad a_3 = -6$$

so by Definition VR [441]

$$\rho_C(\mathbf{u}) = \begin{bmatrix} 5 \\ 16 \\ -6 \end{bmatrix}$$

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 [271] and then arguing that 3 vectors from P_2, a vector space of dimension 3 (Theorem DP [298]), must also be a spanning set (Theorem G [310]). 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,

$$\begin{aligned} 15 &= -2a_1 + a_2 + 5a_3 \\ 10 &= -a_1 + 4a_3 \\ -6 &= 3a_1 - 2a_2 + a_3 \end{aligned}$$

The coefficient matrix of this sytem is nonsingular, leading to a unique solution (no surprise there, see Theorem VRRB [287]),

$$a_1 = -2 \qquad a_2 = 1 \qquad a_3 = 2$$

so by Definition VR [441]

$$\rho_D(\mathbf{u}) = \begin{bmatrix} -2 \\ 1 \\ 2 \end{bmatrix}$$ ⊚

Theorem VRI
Vector Representation is Injective
The function ρ_B (Definition VR [441]) is an injective linear transformation. □

Proof We will appeal to Theorem NSILT [406]. Suppose U is a vector space of dimension n, so vector representation is of the form $\rho_B\colon U \mapsto \mathbb{C}^n$. 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 $\mathbf{u} \in \mathcal{N}(\rho_B)$. Finally, since B is a basis for U, by Theorem VRRB [287] there are (unique) scalars, $a_1,\, a_2,\, a_3,\, \ldots,\, a_n$ such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_n\mathbf{u}_n$$

Then

$$\begin{aligned} \mathbf{0} &= \rho_B\left(\mathbf{u}\right) && \mathbf{u} \in \mathcal{N}(\rho_B) \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} &= \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} \end{aligned}$$

From this vector equality (Definition CVE [90]) we find that $a_i = 0$, $1 \leq i \leq n$. So

$$\begin{aligned} \mathbf{u} &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_n\mathbf{u}_n \\ &= 0\mathbf{u}_1 + 0\mathbf{u}_2 + 0\mathbf{u}_3 + \cdots + 0\mathbf{u}_n \\ &= \mathbf{0} + \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} && \text{Theorem ZSSM [247]} \\ &= \mathbf{0} && \text{Property of } \mathbf{0} \end{aligned}$$

So $\mathcal{N}(\rho_B) = \{\mathbf{0}\}$ and by Theorem NSILT [406], ρ_B is injective. ■

Theorem VRS
Vector Representation is Surjective
The function ρ_B (Definition VR [441]) is a surjective linear transformation. □

Proof We will appeal to Theorem RSLT [419]. Suppose U is a vector space of dimension n, so vector representation is of the form $\rho_B\colon U \mapsto \mathbb{C}^n$. 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 $\mathbf{v} \in \mathbb{C}^n$. Write

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix}$$

and define

$$\mathbf{u} = v_1\mathbf{u}_1 + v_2\mathbf{u}_2 + v_3\mathbf{u}_3 + \cdots + v_n\mathbf{u}_n$$

Then

$$\begin{aligned}\rho_B(\mathbf{u}) &= \rho_B(v_1\mathbf{u}_1 + v_2\mathbf{u}_2 + v_3\mathbf{u}_3 + \cdots + v_n\mathbf{u}_n) \\ &= \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} && \text{Definition VR [441]} \\ &= \mathbf{v}\end{aligned}$$

this demonstrates that $\mathbf{v} \in \mathcal{R}(T)$, so $\mathbb{C}^n \subseteq \mathcal{R}(T)$. Since $\mathcal{R}(T) \subseteq \mathbb{C}^n$ by Definition RLT [416], we have $\mathcal{R}(T) = \mathbb{C}^n$ and Theorem RSLT [419] says T 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 Linear Transformation
The function ρ_B (Definition VR [441]) is an invertible linear transformation. □

Proof The function ρ_B (Definition VR [441]) is a linear transformation (Theorem VRLT [441]) that is injective (Theorem VRI [446]) and surjective (Theorem VRS [447]) with domain V and codomain $\mathbb{C}^n$. By Theorem ILTIS [429] 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 [293]) 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 [432], we have that V and $\mathbb{C}^n$ are isomorphic. ■

Theorem CFDVS [448] 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 [298]). By Theorem CFDVS [448], P_8 is isomorphic to $\mathbb{C}^9$. ⊚

Example CVSR
Crazy vector space revealed
The crazy vector space, C of Example CVS [244], has dimension 2 (can you prove this?). By Theorem CFDVS [448], C is isomorphic to $\mathbb{C}^2$. ⊚

Example ASC
A subspace charaterized
In Example DSP4 [299] we determined that a subspace W of P_4 has dimension 4. By Theorem CFDVS [448], 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 [433].

($\Leftarrow$) This is the advertised converse of Theorem IVSED [433]. 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 [448] there are isomorphisms $T\colon U \mapsto \mathbb{C}^n$ and $S\colon V \mapsto \mathbb{C}^n$.

T is therefore an invertible linear transformation by Definition IVS [432]. Also, S is an invertible linear transformation, so S^{-1} is an invertible linear transformation (Theorem IILT [429]). Invertible linear transformations are injective and surjective (Theorem ILTIS [429]), so the composition of two invertible linear transformations will be injective and surjective (Theorem CILTI [410], Theorem CSLTS [423]), and hence by Theorem ILTIS [429] the composition of S^{-1} with T is invertible.

So $(S^{-1} \circ T) : U \mapsto V$ is an invertible linear transformation from U to V and Definition IVS [432] 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 [449] each is isomorphic to any other.

The subspace of $M_{4,4}$ that contains all the symmetric matrices (Definition SYM [163]) has dimension 10, so this subspace is also isomorphic to each of the four vector spaces above. ⊚

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 [448]). As an invertible linear transformation, ρ_B is an injective linear transformation (Theorem ILTIS [429]), and ρ_B^{-1} is also an injective linear transformation (Theorem IILT [429], Theorem ILTIS [429]).

($\Rightarrow$) Since ρ_B is an injective linear transformation and S is linearly independent, Theorem ILTLI [408] says that R is linearly independent.

($\Leftarrow$) 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 [408] 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 $\mathbf{u} \in \mathcal{S}p(\{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_k\})$ if and only if $\rho_B(\mathbf{u}) \in \mathcal{S}p(\{\rho_B(\mathbf{u}_1),\,\rho_B(\mathbf{u}_2),\,\rho_B(\mathbf{u}_3),\,\ldots,\,\rho_B(\mathbf{u}_k)\})$. □

Proof ($\Rightarrow$) Suppose $\mathbf{u} \in \mathcal{S}p(\{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_k\})$. Then there are scalars, $a_1,\,a_2,\,a_3,\,\ldots,\,a_k$, such that

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_k\mathbf{u}_k$$

Then,

$$\begin{aligned}
\rho_B(\mathbf{u}) &= \rho_B(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_k\mathbf{u}_k)\\
&= a_1\rho_B(\mathbf{u}_1) + a_2\rho_B(\mathbf{u}_2) + a_3\rho_B(\mathbf{u}_3) + \cdots + a_k\rho_B(\mathbf{u}_k) && \text{Theorem LTLC [389]}
\end{aligned}$$

which says that $\rho_B(\mathbf{u}) \in \mathcal{S}p(\{\rho_B(\mathbf{u}_1),\,\rho_B(\mathbf{u}_2),\,\rho_B(\mathbf{u}_3),\,\ldots,\,\rho_B(\mathbf{u}_k)\})$.

($\Leftarrow$) Suppose that $\rho_B(\mathbf{u}) \in \mathcal{S}p(\{\rho_B(\mathbf{u}_1),\,\rho_B(\mathbf{u}_2),\,\rho_B(\mathbf{u}_3),\,\ldots,\,\rho_B(\mathbf{u}_k)\})$. Then there are scalars $b_1,\,b_2,\,b_3,\,\ldots,\,b_k$ such that

$$\rho_B(\mathbf{u}) = b_1\rho_B(\mathbf{u}_1) + b_2\rho_B(\mathbf{u}_2) + b_3\rho_B(\mathbf{u}_3) + \cdots + b_k\rho_B(\mathbf{u}_k)$$

Recall that ρ_B is invertible (Theorem VRILT [448]), so

$$\begin{aligned}
\mathbf{u} &= I_U(\mathbf{u}) && \text{Definition IDLT [425]}\\
&= \left(\rho_B^{-1}\circ\rho_B\right)(\mathbf{u}) && \text{Definition IVLT [425]}\\
&= \rho_B^{-1}(\rho_B(\mathbf{u})) && \text{Definition LTC [397]}\\
&= \rho_B^{-1}(b_1\rho_B(\mathbf{u}_1) + b_2\rho_B(\mathbf{u}_2) + b_3\rho_B(\mathbf{u}_3) + \cdots + b_k\rho_B(\mathbf{u}_k)) && \text{Substitution}\\
&= b_1\rho_B^{-1}(\rho_B(\mathbf{u}_1)) + b_2\rho_B^{-1}(\rho_B(\mathbf{u}_2)) + b_3\rho_B^{-1}(\rho_B(\mathbf{u}_3))\\
&\quad + \cdots + b_k\rho_B^{-1}(\rho_B(\mathbf{u}_k)) && \text{Theorem LTLC [389]}\\
&= b_1 I_U(\mathbf{u}_1) + b_2 I_U(\mathbf{u}_2) + b_3 I_U(\mathbf{u}_3) + \cdots + b_k I_U(\mathbf{u}_k) && \text{Definition IVLT [425]}\\
&= b_1\mathbf{u}_1 + b_2\mathbf{u}_2 + b_3\mathbf{u}_3 + \cdots + b_k\mathbf{u}_k && \text{Definition IDLT [425]}
\end{aligned}$$

which says that $\mathbf{u} \in \mathcal{S}p(\{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_k\})$. ■

Here's a fairly simple example that illustrates a very, very important idea.

Example CP2
Coordinatizing in P_2
In Example VRP2 [445] 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 [449] and Theorem CSS [450] this task is much easier. First, choose a known basis for P_2, a basis that is easy to form vector representations with. 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 CNSMB [285]). Now, since F is a basis for $\mathbb{C}^3$, Theorem CLI [449] and Theorem CSS [450] tell us that D is also a basis for P_2. ⊚

Example CP2 [451] 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 [239] or Chapter LT [379]employing vector spaces of matrices or polynomials. These computations seemed to invariable result in systems of equations or the like from Chapter SLE [3], Chapter V [89] and Chapter M [159]. 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 indepencence 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 [441], Theorem CLI [449] and Theorem CSS [450]. 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 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 (if necessary) through application of the inverse linear transformation ρ_B^{-1}.

Example CM32
Coordinatization in M_{32}
This is a simple example of the Coordinatization Principle [451], depending only on the fact that coordinatizing is an invertible linear transformation (Theorem VRILT [448]). 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 [160], Definition SMM [160]). 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),

$$B = \left\{\begin{bmatrix}1 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix},\ \begin{bmatrix}0 & 0\\ 1 & 0\\ 0 & 0\end{bmatrix},\ \begin{bmatrix}0 & 0\\ 0 & 0\\ 1 & 0\end{bmatrix},\ \begin{bmatrix}0 & 1\\ 0 & 0\\ 0 & 0\end{bmatrix},\ \begin{bmatrix}0 & 0\\ 0 & 1\\ 0 & 0\end{bmatrix},\ \begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & 1\end{bmatrix}\right\}$$

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 [91], Definition CVSM [92]), 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 3×2 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. ⊚

Subsection READ
Reading Questions

1. The vector space of 3×5 matrices, M_{35} 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?

Section MR
Matrix Representations

Definition MR
Matrix Representation
Suppose that $T\colon U \mapsto V$ is a linear transformation, $B = \{\mathbf{u}_1,\,\mathbf{u}_2,\,\mathbf{u}_3,\,\ldots,\,\mathbf{u}_n\}$ is a basis for U of size n, and C is a basis for V of size m. The the **matrix representation** of T relative to B and C is the $m \times n$ matrix,

$$M^T_{B,C} = \left[\rho_C\left(T\left(\mathbf{u}_1\right)\right)\middle|\,\rho_C\left(T\left(\mathbf{u}_2\right)\right)\middle|\,\rho_C\left(T\left(\mathbf{u}_3\right)\right)\middle|\ldots\middle|\rho_C\left(T\left(\mathbf{u}_n\right)\right)\right] \qquad \triangle$$

We may choose to use whatever terms we want when we make a definition. Some are arbitrary, 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 $T\colon U \mapsto V$ is a linear transformation, B is a basis for U, C is a basis for V and $M^T_{B,C}$ is the matrix representation of T relative to B and C. Then, for any $\mathbf{u} \in U$,

$$\rho_C\left(T\left(\mathbf{u}\right)\right) = M^T_{B,C}\left(\rho_B\left(\mathbf{u}\right)\right) \qquad \square$$

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

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_n\mathbf{u}_n$$

Then,

$$\begin{aligned}
&M^T_{B,C}\left(\rho_B\left(\mathbf{u}\right)\right)\\
&= \left[\rho_C\left(T\left(\mathbf{u}_1\right)\right)\middle|\,\rho_C\left(T\left(\mathbf{u}_2\right)\right)\middle|\,\rho_C\left(T\left(\mathbf{u}_3\right)\right)\middle|\ldots\middle|\rho_C\left(T\left(\mathbf{u}_n\right)\right)\right]\mathbf{u} && \text{Definition MR [455]}\\
&= \left[\rho_C\left(T\left(\mathbf{u}_1\right)\right)\middle|\,\rho_C\left(T\left(\mathbf{u}_2\right)\right)\middle|\,\rho_C\left(T\left(\mathbf{u}_3\right)\right)\middle|\ldots\middle|\rho_C\left(T\left(\mathbf{u}_n\right)\right)\right]\begin{bmatrix}a_1\\a_2\\a_3\\\vdots\\a_n\end{bmatrix} && \text{Definition VR [441]}\\
&= a_1\rho_C\left(T\left(\mathbf{u}_1\right)\right) + a_2\rho_C\left(T\left(\mathbf{u}_2\right)\right) + a_3\rho_C\left(T\left(\mathbf{u}_3\right)\right)\\
&\qquad + \cdots + a_n\rho_C\left(T\left(\mathbf{u}_n\right)\right) && \text{Definition MVP [195]}\\
&= \rho_C\left(a_1T\left(\mathbf{u}_1\right) + a_2T\left(\mathbf{u}_2\right) + a_3T\left(\mathbf{u}_3\right) + \cdots + a_nT\left(\mathbf{u}_n\right)\right) && \text{Theorem LTLC [389]}\\
&= \rho_C\left(T\left(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + a_3\mathbf{u}_3 + \cdots + a_n\mathbf{u}_n\right)\right) && \text{Theorem LTLC [389]}\\
&= \rho_C\left(T\left(\mathbf{u}\right)\right) && \text{Substitution} \qquad \blacksquare
\end{aligned}$$

This theorem says that we can apply T to $\mathbf{u}$ and coordinatize the result relative to C in V, or we can first coordinatize $\mathbf{u}$ 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 [195]). That's important enough to say again. The effect of a linear transformation is a matrix-vector product. A minor rearrangement of this result might be even more striking. Apply ρ_C^{-1} to both sides of the conclusion of Theorem FTMR [455] and get

$$T(\mathbf{u}) = \rho_C^{-1}\left(M_{B,C}^{T}\left(\rho_B(\mathbf{u})\right)\right)$$

To effect a linear transformation (T) of a vector ($\mathbf{u}$), coordinatize (ρ_B), do a matrix-vector product ($M_{B,C}^{T}$), and un-coordinatize (ρ_C^{-1}). So, absent some bookkeeping about vector representations, a linear transformation *is* a matrix. Another way to view this is with a diagram similar to those we saw just after defining a linear transformation,

We will use Theorem FTMR [455] 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 [394] we built new linear transformations from other linear transformations. Sums, scalar multiples and compositions. These new linear transformations will have matrix represntations 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 $T\colon U \mapsto V$ and $S\colon U \mapsto V$ 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,

$$\begin{aligned}
M_{B,C}^{T+S}\mathbf{x} &= M_{B,C}^{T+S}\rho_B(\mathbf{u}) && \text{Substitution}\\
&= \rho_C\left((T+S)(\mathbf{u})\right) && \text{Theorem FTMR [455]}\\
&= \rho_C\left(T(\mathbf{u}) + S(\mathbf{u})\right) && \text{Definition LTA [394]}\\
&= \rho_C\left(T(\mathbf{u})\right) + \rho_C\left(S(\mathbf{u})\right) && \text{Definition LT [379]}\\
&= M_{B,C}^{T}\left(\rho_B(\mathbf{u})\right) + M_{B,C}^{S}\left(\rho_B(\mathbf{u})\right) && \text{Theorem FTMR [455]}\\
&= \left(M_{B,C}^{T} + M_{B,C}^{S}\right)\rho_B(\mathbf{u}) && \text{Theorem MMDAA [201]}\\
&= \left(M_{B,C}^{T} + M_{B,C}^{S}\right)\mathbf{x} && \text{Substitution}
\end{aligned}$$

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 $\mathbb{C}^n$, they have equal products on some basis of $\mathbb{C}^n$, and by Theorem XX [**??**], they are equal matrices. ■

Theorem MRMLT
Matrix Representation of a Multiple of a Linear Transformation
Suppose that $T\colon U \mapsto V$ is a linear transformation, $\alpha \in \mathbb{C}$, B is a basis of U and C is a basis 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,

$$\begin{aligned}
M_{B,C}^{\alpha T}\mathbf{x} &= M_{B,C}^{\alpha T}\rho_B(\mathbf{u}) && \text{Substitution}\\
&= \rho_C((\alpha T)(\mathbf{u})) && \text{Theorem FTMR [455]}\\
&= \rho_C(\alpha T(\mathbf{u})) && \text{Definition LTSM [395]}\\
&= \alpha\rho_C(T(\mathbf{u})) && \text{Definition LT [379]}\\
&= \alpha\left(M_{B,C}^{T}\rho_B(\mathbf{u})\right) && \text{Theorem FTMR [455]}\\
&= \left(\alpha M_{B,C}^{T}\right)\rho_B(\mathbf{u}) && \text{Theorem MMSMM [202]}\\
&= \left(\alpha M_{B,C}^{T}\right)\mathbf{x} && \text{Substitution}
\end{aligned}$$

Since the matrices $M_{B,C}^{\alpha T}$ and $\alpha M_{B,C}^{T}$ have equal matrix-vector products for *every* vector in $\mathbb{C}^n$, they have equal products on some basis of $\mathbb{C}^n$, and by Theorem XX [**??**], 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 $T\colon U \mapsto V$ and $S\colon V \mapsto W$ 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,

$$\begin{aligned}
M_{B,D}^{S\circ T}\mathbf{x} &= M_{B,D}^{S\circ T}\rho_B(\mathbf{u}) && \text{Substitution}\\
&= \rho_D((S\circ T)(\mathbf{u})) && \text{Theorem FTMR [455]}\\
&= \rho_D(S(T(\mathbf{u}))) && \text{Definition LTC [397]}\\
&= M_{C,D}^{S}\rho_C(T(\mathbf{u})) && \text{Theorem FTMR [455]}\\
&= M_{C,D}^{S}\left(M_{B,C}^{T}\rho_B(\mathbf{u})\right) && \text{Theorem FTMR [455]}\\
&= \left(M_{C,D}^{S}M_{B,C}^{T}\right)\rho_B(\mathbf{u}) && \text{Theorem MMA [202]}\\
&= \left(M_{C,D}^{S}M_{B,C}^{T}\right)\mathbf{x} && \text{Substitution}
\end{aligned}$$

Since the matrices $M^{S\circ T}_{B,D}$ and $M^{S}_{C,D}M^{T}_{B,C}$ have equal matrix-vector products for *every* vector in $\mathbb{C}^n$, they have equal products on some basis of $\mathbb{C}^n$, and by Theorem XX [**??**], 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 [197]. In either case, we arrive at the same result.

One of our goals in the first part of this book is to make the definition of matrix multiplication (Definition MVP [195], Definition MM [197]) seem as natural as possible. Many are brought up with an entry-by-entry description of matrix multiplication (Theorem ME [360]) 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.

Subsection PMR
Properties of Matrix Representations

It will not be a surprise to discover that the null space and range of a linear transformation are closely related to the null space and range 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 [455]), and a fundamental theorem to go with it (Theorem FTMR [455]) we can be formal about the relationship, using the idea of isomorphic vector spaces (Definition IVS [432]). Here are the twin theorems.

Theorem INS
Isomorphic Null Spaces
Suppose that $T\colon U \mapsto V$ is a linear transformation, B is a basis for U of size n, and C is a basis for V. Then the null space of T is isomorphic to the null space of $M^{T}_{B,C}$,

$$\mathcal{N}(T) \cong \mathcal{N}\left(M^{T}_{B,C}\right)$$ □

Proof To establish that two vector spaces are isomorphic, we must find an isomorphism between them, an invertible linear transformation (Definition IVS [432]). The null space of the linear transformation T, $\mathcal{N}(T)$, is a subspace of U, while the null space of the matrix representation, $\mathcal{N}\left(M^{T}_{B,C}\right)$ 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 $\mathcal{N}(T)$ to $\mathcal{N}\left(M^T_{B,C}\right)$.

The restriction in the size of the domain and codomain will not affect the fact that ρ_B is a linear transformation (Theorem VRLT [441]), nor will it affect the fact that ρ_B is injective (Theorem VRI [446]). Something must done though to verify that ρ_B is surjective. To this end, appeal to the definition of surjective (Definition SLT [411]), and suppose that we have an element of the codomain, $\mathbf{x} \in \mathcal{N}\left(M^T_{B,C}\right) \subseteq \mathbb{C}^n$ 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 $\mathbf{u} \in \mathcal{N}(T)$,

$$
\begin{aligned}
T(\mathbf{u}) &= T\left(\rho_B^{-1}(\mathbf{x})\right) && \text{Substitution}\\
&= I_V\left(T\left(\rho_B^{-1}(\mathbf{x})\right)\right) && \text{Definition IDLT [425]}\\
&= \rho_C^{-1}\left(\rho_C\left(T\left(\rho_B^{-1}(\mathbf{x})\right)\right)\right) && \text{Definition IVLT [425]}\\
&= \rho_C^{-1}\left(M^T_{B,C}\left(\rho_B\left(\rho_B^{-1}(\mathbf{x})\right)\right)\right) && \text{Theorem FTMR [455]}\\
&= \rho_C^{-1}\left(M^T_{B,C}\left(I_{\mathbb{C}^n}(\mathbf{x})\right)\right) && \text{Definition IVLT [425]}\\
&= \rho_C^{-1}\left(M^T_{B,C}\mathbf{x}\right) && \text{Definition IDLT [425]}\\
&= \rho_C^{-1}\left(\mathbf{0}_{\mathbb{C}^n}\right) && \mathbf{x} \in \mathcal{N}\left(M^T_{B,C}\right)\\
&= \mathbf{0}_V && \text{Theorem LTTZZ [383]}
\end{aligned}
$$

Second, verify that the proposed isomorphism, ρ_B, takes $\mathbf{u}$ to $\mathbf{x}$,

$$
\begin{aligned}
\rho_B(\mathbf{u}) &= \rho_B\left(\rho_B^{-1}(\mathbf{x})\right) && \text{Substitution}\\
&= I_{\mathbb{C}^n}(\mathbf{x}) && \text{Definition IVLT [425]}\\
&= \mathbf{x} && \text{Definition IDLT [425]}
\end{aligned}
$$

With ρ_B demonstrated to be an injective and surjective linear transformation from $\mathcal{N}(T)$ to $\mathcal{N}\left(M^T_{B,C}\right)$, Theorem ILTIS [429] tells us ρ_B is invertible, and so by Definition IVS [432], we say $\mathcal{N}(T)$ and $\mathcal{N}\left(M^T_{B,C}\right)$ are isomorphic. ■

Theorem IR
Isomorphic Ranges
Suppose that $T\colon U \mapsto V$ 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 range of $M^T_{B,C}$,

$$\mathcal{R}(T) \cong \mathcal{R}\left(M^T_{B,C}\right)$$ □

Proof To establish that two vector spaces are isomorphic, we must find an isomorphism between them, an invertible linear transformation (Definition IVS [432]). The range of the linear transformation T, $\mathcal{R}(T)$, is a subspace of V, while the range of the matrix representation, $\mathcal{N}\left(M^T_{B,C}\right)$ 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 $\mathcal{R}\left(M^T_{B,C}\right)$.

The restriction in the size of the domain and codomain will not affect the fact that ρ_C is a linear transformation (Theorem VRLT [441]), nor will it affect the fact that ρ_C is injective (Theorem VRI [446]). Something must 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 [411]), and suppose that we have an element of the codomain, $\mathbf{y} \in \mathcal{R}\left(M^T_{B,C}\right) \subseteq \mathbb{C}^m$ and we wish to find an element of the domain with $\mathbf{y}$ as its image. Since $\mathbf{y} \in \mathcal{R}\left(M^T_{B,C}\right)$, there exists a vector, $\mathbf{x} \in \mathbb{C}^n$ with $M^T_{B,C}\mathbf{x} = \mathbf{y}$. 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 $\mathbf{u} = \rho_B^{-1}(\mathbf{x})$,

$$\begin{aligned}
T(\mathbf{u}) &= T\left(\rho_B^{-1}(\mathbf{x})\right) && \text{Substitution}\\
&= I_V\left(T\left(\rho_B^{-1}(\mathbf{x})\right)\right) && \text{Definition IDLT [425]}\\
&= \rho_C^{-1}\left(\rho_C\left(T\left(\rho_B^{-1}(\mathbf{x})\right)\right)\right) && \text{Definition IVLT [425]}\\
&= \rho_C^{-1}\left(M^T_{B,C}\left(\rho_B\left(\rho_B^{-1}(\mathbf{x})\right)\right)\right) && \text{Theorem FTMR [455]}\\
&= \rho_C^{-1}\left(M^T_{B,C}\left(I_{\mathbb{C}^n}(\mathbf{x})\right)\right) && \text{Definition IVLT [425]}\\
&= \rho_C^{-1}\left(M^T_{B,C}\mathbf{x}\right) && \text{Definition IDLT [425]}\\
&= \rho_C^{-1}(\mathbf{y}) && \mathbf{y} \in \mathcal{R}\left(M^T_{B,C}\right)\\
&= \mathbf{v} && \text{Substitution}
\end{aligned}$$

Second, verify that the proposed isomorphism, ρ_C, takes $\mathbf{v}$ to $\mathbf{y}$,

$$\begin{aligned}
\rho_C(\mathbf{v}) &= \rho_C\left(\rho_C^{-1}(\mathbf{y})\right) && \text{Substitution}\\
&= I_{\mathbb{C}^m}(\mathbf{y}) && \text{Definition IVLT [425]}\\
&= \mathbf{y} && \text{Definition IDLT [425]}
\end{aligned}$$

With ρ_C demonstrated to be an injective and surjective linear transformation from $\mathcal{R}(T)$ to $\mathcal{R}\left(M^T_{B,C}\right)$, Theorem ILTIS [429] tells us ρ_C is invertible, and so by Definition IVS [432], we say $\mathcal{R}(T)$ and $\mathcal{R}\left(M^T_{B,C}\right)$ are isomorphic. ■

These two theorems can be viewed as further formal evidence for the Coordinatization Principle [451], 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 transformation that makes this idea precise. Here's our final theorem that solidifies this connection.

Theorem IMR
Invertible Matrix Representations
Suppose that $T\colon U \mapsto V$ is an invertible linear transformation, B is a basis for U and C is a basis for V. Then the matrix representation of T relative to B and C, $M^{T}_{B,C}$ is an invertible matrix, and

$$M^{T^{-1}}_{C,B} = \left(M^{T}_{B,C}\right)^{-1}$$

□

Proof This theorem states that the matrix representation of T^{-1} can be found by finding the matrix inverse of the matrix representation of T (with suitable bases in the right places). It also says that the matrix representation of T is an invertible matrix. We can establish the invertibility, and precisely what the inverse is, by appealing to the definition of a matrix inverse, Definition MI [216]. To this end, let $B = \{\mathbf{u}_1,\, \mathbf{u}_2,\, \mathbf{u}_3,\, \ldots,\, \mathbf{u}_n\}$ and $C = \{\mathbf{v}_1,\, \mathbf{v}_2,\, \mathbf{v}_3,\, \ldots,\, \mathbf{v}_n\}$. Then

$$\begin{aligned}
M^{T^{-1}}_{C,B} M^{T}_{B,C} &= M^{T^{-1}\circ T}_{B,B} && \text{Theorem MRCLT [457]}\\
&= M^{I_U}_{B,B} && \text{Definition IVLT [425]}\\
&= \left[\rho_B\left(I_U\left(\mathbf{u}_1\right)\right)\middle|\, \rho_B\left(I_U\left(\mathbf{u}_2\right)\right)\middle|\ldots\middle|\rho_B\left(I_U\left(\mathbf{u}_n\right)\right)\right] && \text{Definition MR [455]}\\
&= \left[\rho_B\left(\mathbf{u}_1\right)\middle|\, \rho_B\left(\mathbf{u}_2\right)\middle|\, \rho_B\left(\mathbf{u}_3\right)\middle|\ldots\middle|\rho_B\left(\mathbf{u}_n\right)\right] && \text{Definition IDLT [425]}\\
&= \left[\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\ldots|\mathbf{e}_n\right] && \text{Definition VR [441]}\\
&= I_n && \text{Definition IM [76]}
\end{aligned}$$

and

$$\begin{aligned}
M^{T}_{B,C} M^{T^{-1}}_{C,B} &= M^{T\circ T^{-1}}_{C,C} && \text{Theorem MRCLT [457]}\\
&= M^{I_V}_{C,C} && \text{Definition IVLT [425]}\\
&= \left[\rho_C\left(I_V\left(\mathbf{v}_1\right)\right)\middle|\, \rho_C\left(I_V\left(\mathbf{v}_2\right)\right)\middle|\ldots\middle|\rho_C\left(I_V\left(\mathbf{v}_n\right)\right)\right] && \text{Definition MR [455]}\\
&= \left[\rho_C\left(\mathbf{v}_1\right)\middle|\, \rho_C\left(\mathbf{v}_2\right)\middle|\, \rho_C\left(\mathbf{v}_3\right)\middle|\ldots\middle|\rho_C\left(\mathbf{v}_n\right)\right] && \text{Definition IDLT [425]}\\
&= \left[\mathbf{e}_1|\mathbf{e}_2|\mathbf{e}_3|\ldots|\mathbf{e}_n\right] && \text{Definition VR [441]}\\
&= I_n && \text{Definition IM [76]}
\end{aligned}$$

So by Definition MI [216], the matrix $M^{T}_{B,C}$ has an inverse, and that inverse is $M^{T^{-1}}_{C,B}$. ■

THIS SECTION NOT COMPLETE

Subsection READ
Reading Questions

1. Why does Theorem FTMR [455] deserve the moniker “fundamental”?

2. Find the matrix representation, $M^T_{B,C}$ of the linear transformation

$$T\colon \mathbb{C}^2 \mapsto \mathbb{C}^2, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 - x_2 \\ 3x_1 + 2x_2 \end{bmatrix}$$

relative to the bases

$$B = \left\{\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix}\right\} \qquad C = \left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$$

3. What is the second "surprise," and why is it surprising?

Section CB
Change of Basis

We have seen in Section MR [455] 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 $T\colon V \mapsto V$ 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 do exist.

(TODO: Examples here for abstract vector space eigenvectors.)

Subsection CBM
Change-of-Basis Matrix

Definition CBM
Change-of-Basis Matrix
Suppose that V is a vector space, and $I_V\colon V \mapsto V$ 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,

$$\begin{aligned}C_{B,C} &= M^{I_V}_{B,C}\\ &= \left[\rho_C\left(I_V\left(\mathbf{v}_1\right)\right)\middle|\,\rho_C\left(I_V\left(\mathbf{v}_2\right)\right)\middle|\,\rho_C\left(I_V\left(\mathbf{v}_3\right)\right)\middle|\ldots\middle|\rho_C\left(I_V\left(\mathbf{v}_n\right)\right)\right]\\ &= \left[\rho_C\left(\mathbf{u}_1\right)\middle|\,\rho_C\left(\mathbf{u}_2\right)\middle|\,\rho_C\left(\mathbf{u}_3\right)\middle|\ldots\middle|\rho_C\left(\mathbf{u}_n\right)\right]\end{aligned}$$ △

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 $\mathbf{u}$ is a vector in the vector space V and B and C are bases of V. Then

$$C_{B,C}\rho_B\left(\mathbf{v}\right) = \rho_C\left(\mathbf{v}\right)$$ □

Proof

$$\begin{aligned} C_{B,C}\rho_B\left(\mathbf{v}\right) &= M^{I_V}_{B,C}\rho_B\left(\mathbf{v}\right) && \text{Definition CBM [463]}\\ &= \rho_C\left(I_V\left(\mathbf{v}\right)\right) && \text{Theorem FTMR [455]}\\ &= \rho_C\left(\mathbf{v}\right) && \text{Definition IDLT [425]} \end{aligned}$$ ■

So the change-of-basis matrix can be used with matrix multiplication to convert a vector representation of a vector ($\mathbf{v}$) relative to one basis ($\rho_B\left(\mathbf{v}\right)$) to a representation of the same vector relative to a second basis ($\rho_C\left(\mathbf{v}\right)$).

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

$$C^{-1}_{B,C} = C_{C,B}$$ □

Proof The linear transformation $I_V\colon V \mapsto V$ is invertible, and its inverse is itself, I_V (check this!). So by Theorem IMR [461], the matrix $M^{I_V}_{B,C} = C_{B,C}$ is invertible. Theorem NSI [228] says an invertible matrix is nonsingular.

Then

$$\begin{aligned} C^{-1}_{B,C} &= \left(M^{I_V}_{B,C}\right)^{-1} && \text{Definition CBM [463]}\\ &= M^{I_V^{-1}}_{C,B} && \text{Theorem IMR [461]}\\ &= M^{I_V}_{C,B} && \text{Definition IDLT [425]}\\ &= C_{C,B} && \text{Definition CBM [463]} \end{aligned}$$

■

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 $T\colon U \mapsto V$ is a linear transformation, B and C are bases for U, and D and E are bases for V. Then

$$M^{T}_{B,D} = C_{E,D} M^{T}_{C,E} C_{B,C}$$ □

Proof

$$\begin{aligned}
C_{E,D} M^{T}_{C,E} C_{B,C} &= M^{I_V}_{E,D} M^{T}_{C,E} M^{I_U}_{B,C} && \text{Definition CBM [463]}\\
&= M^{I_V}_{E,D} M^{T\circ I_U}_{B,E} && \text{Theorem MRCLT [457]}\\
&= M^{I_V}_{E,D} M^{T}_{B,E} && \text{Definition IDLT [425]}\\
&= M^{I_V\circ T}_{B,D} && \text{Theorem MRCLT [457]}\\
&= M^{T}_{B,D} && \text{Definition IDLT [425]}
\end{aligned}$$ ■

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 $T\colon V \mapsto V$ is a linear transformation and B and C are bases of V. Then

$$M^{T}_{B,B} = C^{-1}_{B,C} M^{T}_{C,C} C_{B,C}$$ □

Proof In the conclusion of Theorem MRCB [465], replace D by B, and replace E by C,

$$\begin{aligned}
M^{T}_{B,B} &= C_{C,B} M^{T}_{C,C} C_{B,C} && \text{Theorem MRCB [465]}\\
&= C^{-1}_{B,C} M^{T}_{C,C} C_{B,C} && \text{Theorem ICBM [464]}
\end{aligned}$$ ■

This is the third surprise of this chapter. Theorem SCB [465] 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 ($M^{T}_{B,B}$), and then we build a second matrix representation of T, now using the basis C for both the domain and codomain ($M^{T}_{C,C}$). Then these two representations are related via a similarity transformation (Definition SIM [365]) using a change-of-basis matrix ($C_{B,C}$)!

We can now return to the question of computing an eigenvalue or eigenvector of a linear transformation. For a linear transformation of the form $T\colon V \mapsto V$, we know that representations relative to different bases are similar matrices. We also know that similar matrices have equal characteristic polynomials by Theorem SMEE [368]. 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 [331].

Theorem EER
Eigenvalues, Eigenvectors, Representations
Suppose that $T\colon V \mapsto V$ 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 $M^T_{B,B}$ for the eigenvalue λ. □

Proof (⇒) Assume that $\mathbf{v} \in V$ is an eigenvector of T for the eigenvalue λ. Then

$$\begin{aligned} M^T_{B,B}\rho_B(\mathbf{v}) &= \rho_B(T(\mathbf{v})) && \text{Theorem FTMR [455]}\\ &= \rho_B(\lambda\mathbf{v}) && \text{Hypothesis}\\ &= \lambda\rho_B(\mathbf{v}) && \text{Theorem VRLT [441]} \end{aligned}$$

which by Definition EEM [331] says that $\rho_B(\mathbf{v})$ is an eigenvector of the matrix $M^T_{B,B}$ for the eigenvalue λ.

(⇐) Assume that $\rho_B(\mathbf{v})$ is an eigenvector of $M^T_{B,B}$ for the eigenvalue λ. Then

$$\begin{aligned} T(\mathbf{v}) &= \rho_B^{-1}(\rho_B(T(\mathbf{v}))) && \text{Definition IVLT [425]}\\ &= \rho_B^{-1}\left(M^T_{B,B}\rho_B(\mathbf{v})\right) && \text{Theorem FTMR [455]}\\ &= \rho_B^{-1}(\lambda\rho_B(\mathbf{v})) && \text{Hypothesis}\\ &= \lambda\rho_B^{-1}(\rho_B(\mathbf{v})) && \text{Theorem ILTLT [428]}\\ &= \lambda\mathbf{v} && \text{Definition IVLT [425]} \end{aligned}$$

which by Definition EELT [463] says $\mathbf{v}$ is an eigenvector of T for the eigenvalue λ. ■

Knowing that the eigenvalues of a linear transformation are the eigenvalues of any representation, no matter what the choice of the basis is, we could now unambigously define items such as the characteristic polynomial of a linear transformation. But we won't go to the trouble.

As a practical matter, how does one compute the eigenvalues and eigenvectors of a linear transformation of the form $T\colon V \mapsto V$? 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 [331]. 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 indepenedent eigenvectors that must exist for the matrix (Theorem DC [370]) 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 [365] were really finding especially pleasing matrix representations of linear transformations.

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 $\mathbb{C}^2$

$$B = \left\{ \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\} \qquad C = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

3. What is the third "surprise," and why is it surprising?

A: Archetypes

The American Heritage Dictionary of the English Language (Third Edition) gives two definitions of the word “archetype”: 1. An original model or type after which other similar things are patterned; a prototype; and 2. An ideal example of a type; quintessence.

Either use might apply here. 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 [473] and Archetype B [478]. Some we have left for you to investigate, such as Archetype J [515], which parallels Archetype I [510].

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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Type	S	S	S	S	S	S	S	S	S	S	M	M	L	L	L	L	L	L
Vars, Cols, Domain	3	3	4	4	4	4	2	2	7	9	5	5	5	5	3	3	5	5
Eqns, Rows, CoDom	3	3	3	3	3	4	5	5	4	6	5	5	3	3	5	5	5	5
Consistent	I	U	I	I	N	U	U	N	I	I								
Rank	2	3	3	2	2	4	2	2	3	4	5	3	2	3	2	3	4	5
Nullity	1	0	1	2	2	0	3	3	4	5	0	2	3	2	1	0	1	0
Injective													X	X	N	Y	N	Y
Surjective													N	Y	X	X	N	Y
Full Rank	N	Y	Y	N	N	Y	Y	Y	N	N	Y	N						
Nonsingular	N	Y				Y					Y	N						
Invertible	N	Y				Y					Y	N					N	Y
Determinant	0	-2				-18					16	0						
Diagonalizable	N	Y				Y					Y	Y						

Archetype Facts
S=System of Equations, M=Matrix, L=Linear Transformation
U=Unique solution, I=Infinitely many solutions, N=No solutions
Y=Yes, N=No, X=Impossible, blank=Not Applicable

Archetype A

Summary Linear system of three equations, three unknowns. Singular coefficent matrix with dimension 1 null space. Integer eigenvalues and a degenerate eigenspace for coefficient matrix.

□ A system of linear equations (Definition SSLE [16]):

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 + x_2 + x_3 &= 8 \\ x_1 + x_2 &= 5 \end{aligned}$$

□ 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 [34]):

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

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3,\, 4\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 0\\ 2x_1 + x_2 + x_3 &= 0\\ x_1 + x_2 \qquad &= 0 \end{aligned}$$

□ 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:

$$\begin{bmatrix}\boxed{1} & 0 & 1 & 0\\ 0 & \boxed{1} & -1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2 \qquad D = \{1, 2\} \qquad F = \{3, 4\}$$

□ 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:

$$\begin{bmatrix} \boxed{1} & 0 & 1 \\ 0 & \boxed{1} & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3\}$$

□ Matrix (coefficient matrix) is nonsingular or singular? (Theorem NSRRI [76]) 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix},\ \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

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

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-3\\0\\1\end{bmatrix},\,\begin{bmatrix}2\\1\\0\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\-\frac{1}{3}\end{bmatrix},\,\begin{bmatrix}0\\1\\\frac{2}{3}\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\1\end{bmatrix},\,\begin{bmatrix}0\\1\\-1\end{bmatrix}\right\}\right)$$

□ 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 [216], Theorem NSI [228])

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 [328]). (Product of all eigenvalues?)

Determinant $=$ 0

☐ Eigenvalues, and bases for eigenspaces. (Definition EEM [331],Definition EM [341])

$$\lambda = 0 \qquad E_A(0) = \mathcal{S}p\left(\left\{\begin{bmatrix}-1\\1\\1\end{bmatrix}\right\}\right)$$

$$\lambda = 2 \qquad E_A(2) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\5\\3\end{bmatrix}\right\}\right)$$

☐ Geometric and algebraic multiplicities. (Definition GME [344]Definition AME [343])

$$\gamma_A(0) = 1 \qquad \alpha_A(0) = 2$$
$$\gamma_A(2) = 1 \qquad \alpha_A(2) = 1$$

☐ Diagonalizable? (Definition DZM [369])

No, $\gamma_A(0) \neq \alpha_B(0)$, Theorem DMLE [373].

Archetype B

Summary System with three equations, three unknowns. Nonsingular coefficent matrix. Distinct integer eigenvalues for coefficient matrix.

□ A system of linear equations (Definition SSLE [16]):

$$\begin{aligned} -7x_1 - 6x_2 - 12x_3 &= -33 \\ 5x_1 + 5x_2 + 7x_3 &= 24 \\ x_1 + 4x_3 &= 5 \end{aligned}$$

□ 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 [34]):

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

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\, 2,\, 3\} \qquad F = \{4\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} -11x_1 + 2x_2 - 14x_3 &= 0 \\ 23x_1 - 6x_2 + 33x_3 &= 0 \\ 14x_1 - 2x_2 + 17x_3 &= 0 \end{aligned}$$

☐ 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:

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \end{bmatrix}$$

☐ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 3 \qquad D = \{1,\, 2,\, 3\} \qquad F = \{4\}$$

☐ 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:

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \end{bmatrix}$$

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\,2,\,3\} \qquad F = \{\ \}$$

□ Matrix (coefficient matrix) is nonsingular or singular? (Theorem NSRRI [76]) 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$\mathcal{S}p(\{\ \})$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-7\\5\\1\end{bmatrix},\,\begin{bmatrix}-6\\5\\0\end{bmatrix},\,\begin{bmatrix}-12\\7\\4\end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$K = []$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$

☐ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$

☐ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$

☐ 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 [216], Theorem NSI [228])

$$\begin{bmatrix}-10 & -12 & -9\\ \frac{13}{2} & 8 & \frac{11}{2}\\ \frac{5}{2} & 3 & \frac{5}{2}\end{bmatrix}$$

☐ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 [328]). (Product of all eigenvalues?)

Determinant $= -2$

☐ Eigenvalues, and bases for eigenspaces. (Definition EEM [331],Definition EM [341])

$$\lambda = -1 \qquad E_B(-1) = \mathcal{S}p\left(\left\{\begin{bmatrix}-5\\3\\1\end{bmatrix}\right\}\right)$$

$$\lambda = 1 \qquad E_B(1) = \mathcal{S}p\left(\left\{\begin{bmatrix}-3\\2\\1\end{bmatrix}\right\}\right)$$

$$\lambda = 2 \qquad E_B(2) = \mathcal{S}p\left(\left\{\begin{bmatrix}-2\\1\\1\end{bmatrix}\right\}\right)$$

□ Geometric and algebraic multiplicities. (Definition GME [344]Definition AME [343])

$$\gamma_B(-1) = 1 \qquad \alpha_B(-1) = 1$$
$$\gamma_B(1) = 1 \qquad \alpha_B(1) = 1$$
$$\gamma_B(2) = 1 \qquad \alpha_B(2) = 1$$

□ Diagonalizable? (Definition DZM [369])

Yes, distinct eigenvalues, Theorem DED [375].

□ The diagonalization. (Theorem DC [370])

$$\begin{bmatrix}-1 & -1 & -1\\ 2 & 3 & 1\\ -1 & -2 & 1\end{bmatrix}\begin{bmatrix}-7 & -6 & -12\\ 5 & 5 & 7\\ 1 & 0 & 4\end{bmatrix}\begin{bmatrix}-5 & -3 & -2\\ 3 & 2 & 1\\ 1 & 1 & 1\end{bmatrix}$$
$$= \begin{bmatrix}-1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 2\end{bmatrix}$$

Archetype C

Summary System with three equations, four variables. Consistent. Null space of coefficient matrix has dimension 1.

□ A system of linear equations (Definition SSLE [16]):

$$\begin{aligned} 2x_1 - 3x_2 + x_3 - 6x_4 &= -7 \\ 4x_1 + x_2 + 2x_3 + 9x_4 &= -7 \\ 3x_1 + x_2 + x_3 + 8x_4 &= -8 \end{aligned}$$

□ 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 [34]):

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

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\, 2,\, 3\} \qquad F = \{4,\, 5\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} 2x_1 - 3x_2 + x_3 - 6x_4 &= 0 \\ 4x_1 + x_2 + 2x_3 + 9x_4 &= 0 \\ 3x_1 + x_2 + x_3 + 8x_4 &= 0 \end{aligned}$$

□ Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

$x_1 = 0, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$

$x_1 = -2, \quad x_2 = -3, \quad x_3 = 1, \quad x_4 = 1$

$x_1 = -4, \quad x_2 = -6, \quad x_3 = 2, \quad x_4 = 2$

□ 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:

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

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 3 \qquad D = \{1,\, 2,\, 3\} \qquad F = \{4,\, 5\}$$

□ 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:

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

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\, 2,\, 3\} \qquad F = \{4\}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix} -2 \\ -3 \\ 1 \\ 1 \end{bmatrix}\right\}\right)$$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix},\, \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix},\, \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$K = [\,]$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\2\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\3\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\-1\end{bmatrix}\right\}\right)$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 SSLE [16]):

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 8 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= -12 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 4 \end{aligned}$$

□ Some solutions to the system of linear equations (not necessarily exhaustive):

$x_1 = 0, \quad x_2 = 1, \quad x_3 = 2, \quad x_4 = 1$

$x_1 = 4, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$

$x_1 = 7, \quad x_2 = 8, \quad x_3 = 1, \quad x_4 = 3$

□ Augmented matrix of the linear system of equations (Definition AM [34]):

$$\begin{bmatrix} 2 & 1 & 7 & -7 & 8 \\ -3 & 4 & -5 & -6 & -12 \\ 1 & 1 & 4 & -5 & 4 \end{bmatrix}$$

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3,\, 4,\, 5\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 0 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 0 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 0 \end{aligned}$$

□ Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):

$x_1 = 0, \quad x_2 = 0, \quad x_3 = 0, \quad x_4 = 0$

$x_1 = -3, \quad x_2 = -1, \quad x_3 = 1, \quad x_4 = 0$

$x_1 = 2, \quad x_2 = 3, \quad x_3 = 0, \quad x_4 = 1$

$x_1 = -1, \quad x_2 = 2, \quad x_3 = 1, \quad x_4 = 1$

□ 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:

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

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3,\, 4,\, 5\}$$

□ 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:

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

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3,\, 4\}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix} -3 \\ -1 \\ 1 \\ 0 \end{bmatrix},\, \begin{bmatrix} 2 \\ 3 \\ 0 \\ 1 \end{bmatrix}\right\}\right)$$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix},\, \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the

range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix} 1 & \frac{1}{7} & -\frac{11}{7} \end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix} \frac{11}{7} \\ 0 \\ 1 \end{bmatrix},\ \begin{bmatrix} -\frac{1}{7} \\ 1 \\ 0 \end{bmatrix}\right\}\right)$$

☐ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 0 \\ \frac{7}{11} \end{bmatrix},\ \begin{bmatrix} 0 \\ 1 \\ \frac{1}{11} \end{bmatrix}\right\}\right)$$

☐ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 0 \\ 3 \\ -2 \end{bmatrix},\ \begin{bmatrix} 0 \\ 1 \\ 1 \\ -3 \end{bmatrix}\right\}\right)$$

☐ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 SSLE [16]):

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 2 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 3 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 2 \end{aligned}$$

☐ Some solutions to the system of linear equations (not necessarily exhaustive):

None. (Why?)

☐ Augmented matrix of the linear system of equations (Definition AM [34]):

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

☐ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

☐ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\, 2,\, 5\} \qquad F = \{3,\, 4\}$$

☐ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 0 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 0 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 0 \end{aligned}$$

□ 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:

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

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2 \qquad D = \{1, 2\} \qquad F = \{3, 4, 5\}$$

□ 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:

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

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3,\, 4\}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix},\,\begin{bmatrix}2\\3\\0\\1\end{bmatrix}\right\}\right)$$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}2\\-3\\1\end{bmatrix},\,\begin{bmatrix}1\\4\\1\end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix} 1 & \frac{1}{7} & -\frac{11}{7} \end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}\frac{11}{7}\\0\\1\end{bmatrix},\ \begin{bmatrix}-\frac{1}{7}\\1\\0\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\\frac{7}{11}\end{bmatrix},\ \begin{bmatrix}0\\1\\\frac{1}{11}\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\3\\-2\end{bmatrix},\ \begin{bmatrix}0\\1\\1\\-3\end{bmatrix}\right\}\right)$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 SSLE [16]):

$$\begin{aligned} 33x_1 - 16x_2 + 10x_3 - 2x_4 &= -27 \\ 99x_1 - 47x_2 + 27x_3 - 7x_4 &= -77 \\ 78x_1 - 36x_2 + 17x_3 - 6x_4 &= -52 \\ -9x_1 + 2x_2 + 3x_3 + 4x_4 &= 5 \end{aligned}$$

□ 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 [34]):

$$\begin{bmatrix} 33 & -16 & 10 & -2 & -27 \\ 99 & -47 & 27 & -7 & -77 \\ 78 & -36 & 17 & -6 & -52 \\ -9 & 2 & 3 & 4 & 5 \end{bmatrix}$$

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 4 \qquad D = \{1,\, 2,\, 3,\, 4\} \qquad F = \{5\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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.

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}1\\2\\-2\\4\end{bmatrix}$$

☐ Given a system of equations we can always build a new, related, homogeneous system (Definition HS [67]) 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.

$$\begin{aligned}33x_1-16x_2+10x_3-2x_4&=0\\99x_1-47x_2+27x_3-7x_4&=0\\78x_1-36x_2+17x_3-6x_4&=0\\-9x_1+2x_2+3x_3+4x_4&=0\end{aligned}$$

☐ 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:

$$\begin{bmatrix}\boxed{1}&0&0&0&0\\0&\boxed{1}&0&0&0\\0&0&\boxed{1}&0&0\\0&0&0&\boxed{1}&0\end{bmatrix}$$

☐ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r=4\qquad D=\{1,\,2,\,3,\,4\}\qquad F=\{5\}$$

☐ 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:

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

☐ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 4 \qquad D = \{1,\ 2,\ 3,\ 4\} \qquad F = \{\ \}$$

☐ Matrix (coefficient matrix) is nonsingular or singular? (Theorem NSRRI [76]) 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$\mathcal{S}p(\{\ \})$

☐ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}33\\99\\78\\-9\end{bmatrix},\,\begin{bmatrix}-16\\-47\\-36\\2\end{bmatrix},\,\begin{bmatrix}10\\27\\17\\3\end{bmatrix},\,\begin{bmatrix}-2\\-7\\-6\\4\end{bmatrix}\right\}\right)$$

☐ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = [\,]$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}\right)$$

☐ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}\right)$$

☐ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}\right)$$

☐ 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 [216], Theorem NSI [228])

$$\begin{bmatrix} -\left(\frac{86}{3}\right) & \frac{38}{3} & -\left(\frac{11}{3}\right) & \frac{7}{3} \\ -\left(\frac{129}{2}\right) & \frac{86}{3} & -\left(\frac{17}{2}\right) & \frac{31}{6} \\ -13 & 6 & -2 & 1 \\ -\left(\frac{45}{2}\right) & \frac{29}{3} & -\left(\frac{5}{2}\right) & \frac{13}{6} \end{bmatrix}$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 [328]). (Product of all eigenvalues?)

Determinant = −18

□ Eigenvalues, and bases for eigenspaces. (Definition EEM [331],Definition EM [341])

$$\lambda = -1 \qquad E_F(-1) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\0\\1\end{bmatrix}\right\}\right)$$

$$\lambda = 2 \qquad E_F(2) = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\5\\2\\1\end{bmatrix}\right\}\right)$$

$$\lambda = 3 \qquad E_F(3) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\1\\0\\7\end{bmatrix}, \begin{bmatrix}17\\45\\21\\0\end{bmatrix}\right\}\right)$$

□ Geometric and algebraic multiplicities. (Definition GME [344]Definition AME [343])

$$\gamma_F(-1) = 1 \qquad \alpha_F(-1) = 1$$
$$\gamma_F(2) = 1 \qquad \alpha_F(2) = 1$$
$$\gamma_F(3) = 2 \qquad \alpha_F(3) = 2$$

□ Diagonalizable? (Definition DZM [369])

Yes, large eigenspaces, Theorem DMLE [373].

☐ The diagonalization. (Theorem DC [370])

$$\begin{bmatrix} 12 & -5 & 1 & -1 \\ -39 & 18 & -7 & 3 \\ \frac{27}{7} & -\frac{13}{7} & \frac{6}{7} & -\frac{1}{7} \\ \frac{26}{7} & -\frac{12}{7} & \frac{5}{7} & -\frac{2}{7} \end{bmatrix} \begin{bmatrix} 33 & -16 & 10 & -2 \\ 99 & -47 & 27 & -7 \\ 78 & -36 & 17 & -6 \\ -9 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 17 \\ 2 & 5 & 1 & 45 \\ 0 & 2 & 0 & 21 \\ 1 & 1 & 7 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

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 SSLE [16]):

$$\begin{aligned} 2x_1 + 3x_2 &= 6 \\ -x_1 + 4x_2 &= -14 \\ 3x_1 + 10x_2 &= -2 \\ 3x_1 - x_2 &= 20 \\ 6x_1 + 9x_2 &= 18 \end{aligned}$$

□ 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 [34]):

$$\begin{bmatrix} 2 & 3 & 6 \\ -1 & 4 & -14 \\ 3 & 10 & -2 \\ 3 & -1 & 20 \\ 6 & 9 & 18 \end{bmatrix}$$

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} 2x_1 + 3x_2 &= 0 \\ -x_1 + 4x_2 &= 0 \\ 3x_1 + 10x_2 &= 0 \\ 3x_1 - x_2 &= 0 \\ 6x_1 + 9x_2 &= 0 \end{aligned}$$

□ 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:

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3\}$$

□ 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.

$$\begin{bmatrix} 2 & 3 \\ -1 & 4 \\ 3 & 10 \\ 3 & -1 \\ 6 & 9 \end{bmatrix}$$

□ Matrix brought to reduced row-echelon form:

$$\begin{bmatrix} \boxed{1} & 0 \\ 0 & \boxed{1} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{\ \}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$\mathcal{S}p(\{\ \})$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 2 \\ -1 \\ 3 \\ 3 \\ 6 \end{bmatrix},\, \begin{bmatrix} 3 \\ 4 \\ 10 \\ -1 \\ 9 \end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the

range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix} 1 & 0 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 1 - \frac{1}{3} & \\ 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \\ 1 \\ 0 \end{bmatrix}\right\}\right)$$

☐ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}\right\}\right)$$

☐ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}\right)$$

☐ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 SSLE [16]):

$$\begin{aligned} 2x_1 + 3x_2 &= 5 \\ -x_1 + 4x_2 &= 6 \\ 3x_1 + 10x_2 &= 2 \\ 3x_1 - x_2 &= -1 \\ 6x_1 + 9x_2 &= 3 \end{aligned}$$

□ Some solutions to the system of linear equations (not necessarily exhaustive):

None. (Why?)

□ Augmented matrix of the linear system of equations (Definition AM [34]):

$$\begin{bmatrix} 2 & 3 & 5 \\ -1 & 4 & 6 \\ 3 & 10 & 2 \\ 3 & -1 & -1 \\ 6 & 9 & 3 \end{bmatrix}$$

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & \boxed{1} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\,2,\,3\} \qquad F = \{\ \}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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 [67]) 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.

$$\begin{aligned} 2x_1 + 3x_2 &= 0 \\ -x_1 + 4x_2 &= 0 \\ 3x_1 + 10x_2 &= 0 \\ 3x_1 - x_2 &= 0 \\ 6x_1 + 9x_2 &= 0 \end{aligned}$$

□ 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:

$$\begin{bmatrix} \boxed{1} & 0 & 0 \\ 0 & \boxed{1} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{3\}$$

□ 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.

$$\begin{bmatrix} 2 & 3 \\ -1 & 4 \\ 3 & 10 \\ 3 & -1 \\ 6 & 9 \end{bmatrix}$$

□ Matrix brought to reduced row-echelon form:

$$\begin{bmatrix} \boxed{1} & 0 \\ 0 & \boxed{1} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 2 \qquad D = \{1,\, 2\} \qquad F = \{\ \}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$\mathcal{S}p(\{\ \})$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix} 2 \\ -1 \\ 3 \\ 3 \\ 6 \end{bmatrix},\, \begin{bmatrix} 3 \\ 4 \\ 10 \\ -1 \\ 9 \end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the

range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = []$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}\frac{1}{3}\\\frac{1}{3}\\1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\-1\\-1\\1\\0\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\2\\1\\3\end{bmatrix}, \begin{bmatrix}0\\1\\1\\-1\\0\end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix}1 & 0 & 0 & 0 & -\frac{1}{3}\\0 & 1 & 0 & 1-\frac{1}{3} & \\0 & 0 & 1 & 1 & -1\end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}\frac{1}{3}\\\frac{1}{3}\\1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\-1\\-1\\1\\0\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\end{bmatrix},\,\begin{bmatrix}0\\1\end{bmatrix}\right\}\right)$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 SSLE [16]):

$$\begin{aligned} x_1 + 4x_2 - x_4 + 7x_6 - 9x_7 &= 3 \\ 2x_1 + 8x_2 - x_3 + 3x_4 + 9x_5 - 13x_6 + 7x_7 &= 9 \\ 2x_3 - 3x_4 - 4x_5 + 12x_6 - 8x_7 &= 1 \\ -x_1 - 4x_2 + 2x_3 + 4x_4 + 8x_5 - 31x_6 + 37x_7 &= 4 \end{aligned}$$

□ 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 [34]):

$$\begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 & 3 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 & 9 \\ 0 & 0 & 2 & -3 & -4 & 12 & -8 & 1 \\ -1 & -4 & 2 & 4 & 8 & -31 & 37 & 4 \end{bmatrix}$$

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

□ Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\ 3,\ 4\} \qquad F = \{2,\ 5,\ 6,\ 7,\ 8\}$$

□ Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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.

$$\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\end{bmatrix}=\begin{bmatrix}4\\0\\2\\1\\0\\0\\0\end{bmatrix}+x_2\begin{bmatrix}-4\\1\\0\\0\\0\\0\\0\end{bmatrix}+x_5\begin{bmatrix}-2\\0\\-1\\-2\\1\\0\\0\end{bmatrix}+x_6\begin{bmatrix}-1\\0\\3\\6\\0\\1\\0\end{bmatrix}+x_7\begin{bmatrix}3\\0\\-5\\-6\\0\\0\\1\end{bmatrix}$$

☐ Given a system of equations we can always build a new, related, homogeneous system (Definition HS [67]) 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.

$$\begin{aligned}x_1+4x_2-x_4+7x_6-9x_7&=0\\2x_1+8x_2-x_3+3x_4+9x_5-13x_6+7x_7&=0\\2x_3-3x_4-4x_5+12x_6-8x_7&=0\\-x_1-4x_2+2x_3+4x_4+8x_5-31x_6+37x_7&=0\end{aligned}$$

☐ 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:

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

□ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 3 \qquad D = \{1,\,3,\,4\} \qquad F = \{2,\,5,\,6,\,7,\,8\}$$

□ 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.

$$\begin{bmatrix} 1 & 4 & 0 & -1 & 0 & 7 & -9 \\ 2 & 8 & -1 & 3 & 9 & -13 & 7 \\ 0 & 0 & 2 & -3 & -4 & 12 & -8 \\ -1 & -4 & 2 & 4 & 8 & -31 & 37 \end{bmatrix}$$

□ Matrix brought to reduced row-echelon form:

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

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 3 \qquad D = \{1,\,3,\,4\} \qquad F = \{2,\,5,\,6,\,7\}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-4\\1\\0\\0\\0\\0\\0\end{bmatrix},\begin{bmatrix}-2\\0\\-1\\-2\\1\\0\\0\end{bmatrix},\begin{bmatrix}-1\\0\\3\\6\\0\\1\\0\end{bmatrix},\begin{bmatrix}3\\0\\-5\\-6\\0\\0\\1\end{bmatrix}\right\}\right)$$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\0\\-1\end{bmatrix},\begin{bmatrix}0\\-1\\2\\2\end{bmatrix},\begin{bmatrix}-1\\3\\-3\\4\end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix}1 & -\frac{12}{31} & -\frac{13}{31} & \frac{7}{31}\end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-\frac{7}{31}\\0\\0\\1\end{bmatrix},\begin{bmatrix}\frac{13}{31}\\0\\1\\0\end{bmatrix},\begin{bmatrix}\frac{12}{31}\\1\\0\\0\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\-\frac{31}{7}\end{bmatrix},\begin{bmatrix}0\\1\\0\\\frac{12}{7}\end{bmatrix},\begin{bmatrix}0\\0\\1\\\frac{13}{7}\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\4\\0\\0\\2\\1\\-3\end{bmatrix},\ \begin{bmatrix}0\\0\\1\\0\\1\\-3\\5\end{bmatrix},\ \begin{bmatrix}0\\0\\0\\1\\2\\-6\\6\end{bmatrix}\right\}\right)$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 SSLE [16]):

$$\begin{aligned}
x_1 + 2x_2 - 2x_3 + 9x_4 + 3x_5 - 5x_6 - 2x_7 + x_8 + 27x_9 &= -5\\
2x_1 + 4x_2 + 3x_3 + 4x_4 - x_5 + 4x_6 + 10x_7 + 2x_8 - 23x_9 &= 18\\
x_1 + 2x_2 + x_3 + 3x_4 + x_5 + x_6 + 5x_7 + 2x_8 - 7x_9 &= 6\\
2x_1 + 4x_2 + 3x_3 + 4x_4 - 7x_5 + 2x_6 + 4x_7 - 11x_9 &= 20\\
x_1 + 2x_2 + +5x_4 + 2x_5 - 4x_6 + 3x_7 + 8x_8 + 13x_9 &= -4\\
-3x_1 - 6x_2 - x_3 - 13x_4 + 2x_5 - 5x_6 - 4x_7 + 13x_8 + 10x_9 &= -29
\end{aligned}$$

□ 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 [34]):

$$\begin{bmatrix}
1 & 2 & -2 & 9 & 3 & -5 & -2 & 1 & 27 & -5\\
2 & 4 & 3 & 4 & -1 & 4 & 10 & 2 & -23 & 18\\
1 & 2 & 1 & 3 & 1 & 1 & 5 & 2 & -7 & 6\\
2 & 4 & 3 & 4 & -7 & 2 & 4 & 0 & -11 & 20\\
1 & 2 & 0 & 5 & 2 & -4 & 3 & 8 & 13 & -4\\
-3 & -6 & -1 & -13 & 2 & -5 & -4 & 13 & 10 & -29
\end{bmatrix}$$

□ Matrix in reduced row-echelon form, row-equivalent to augmented matrix:

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

Analysis of the augmented matrix (Notation RREFA [51]):

$$r = 4 \qquad D = \{1,\, 3,\, 5,\, 6\} \qquad F = \{2,\, 4,\, 7,\, 8,\, 9,\, 10\}$$

Vector form of the solution set to the system of equations (Theorem VFSLS [104]). 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.

$$\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6\\ x_7\\ x_8\\ x_9 \end{bmatrix} = \begin{bmatrix} 6\\ 0\\ -1\\ 0\\ -1\\ 2\\ 0\\ 0\\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -2\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -5\\ 0\\ 2\\ 1\\ 0\\ 0\\ 0\\ 0\\ 0 \end{bmatrix} + x_7 \begin{bmatrix} -1\\ 0\\ -3\\ 0\\ -1\\ 0\\ 1\\ 0\\ 0 \end{bmatrix} + x_8 \begin{bmatrix} 2\\ 0\\ -5\\ 0\\ -1\\ 2\\ 0\\ 1\\ 0 \end{bmatrix} + x_9 \begin{bmatrix} -3\\ 0\\ 6\\ 0\\ 1\\ 3\\ 0\\ 0\\ 1 \end{bmatrix}$$

Given a system of equations we can always build a new, related, homogeneous system (Definition HS [67]) 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.

$$\begin{aligned} x_1 + 2x_2 - 2x_3 + 9x_4 + 3x_5 - 5x_6 - 2x_7 + x_8 + 27x_9 &= 0\\ 2x_1 + 4x_2 + 3x_3 + 4x_4 - x_5 + 4x_6 + 10x_7 + 2x_8 - 23x_9 &= 0\\ x_1 + 2x_2 + x_3 + 3x_4 + x_5 + x_6 + 5x_7 + 2x_8 - 7x_9 &= 0\\ 2x_1 + 4x_2 + 3x_3 + 4x_4 - 7x_5 + 2x_6 + 4x_7 - 11x_9 &= 0\\ x_1 + 2x_2 + +5x_4 + 2x_5 - 4x_6 + 3x_7 + 8x_8 + 13x_9 &= 0\\ -3x_1 - 6x_2 - x_3 - 13x_4 + 2x_5 - 5x_6 - 4x_7 + 13x_8 + 10x_9 &= 0 \end{aligned}$$

Some solutions to the associated homogenous system of linear equations (not necessarily exhaustive):
$x_1 = 0,\ x_2 = 0,\ x_3 = 0,\ x_4 = 0,\ x_5 = 0,\ x_6 = 0,\ x_7 = 0,\ x_8 = 0,\ x_9 = 0$

$x_1 = -2,\ x_2 = 1,\ x_3 = 0,\ x_4 = 0,\ x_5 = 0,\ x_6 = 0,\ x_7 = 0,\ x_8 = 0,\ x_9 = 0$

$x_1 = -23,\ x_2 = 7,\ x_3 = 4,\ x_4 = 2,\ x_5 = 0,\ x_6 = 12,\ x_7 = -1,\ x_8 = 3,\ x_9 = 2$

$x_1 = -17,\ x_2 = -6,\ x_3 = 2,\ x_4 = 5,\ x_5 = -3,\ x_6 = 5,\ x_7 = 3,\ x_8 = 1,\ x_9 = 1$

☐ 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:

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

☐ Analysis of the augmented matrix for the homogenous system (Notation RREFA [51]). Notice the slight variation for the same analysis of the original system only when the original system was consistent:

$$r = 4 \qquad D = \{1,\ 3,\ 5,\ 6\} \qquad F = \{2,\ 4,\ 7,\ 8,\ 9,\ 10\}$$

☐ 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:

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

☐ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 4 \qquad D = \{1,\ 3,\ 5,\ 6\} \qquad F = \{2,\ 4,\ 7,\ 8,\ 9\}$$

□ 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-2\\1\\0\\0\\0\\0\\0\\0\\0\end{bmatrix},\begin{bmatrix}-5\\0\\2\\1\\0\\0\\0\\0\\0\end{bmatrix},\begin{bmatrix}-1\\0\\-3\\0\\-1\\0\\1\\0\\0\end{bmatrix},\begin{bmatrix}2\\0\\-5\\0\\-1\\2\\0\\1\\0\end{bmatrix},\begin{bmatrix}-3\\0\\6\\0\\1\\3\\0\\0\\1\end{bmatrix}\right\}\right)$$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\1\\2\\1\\-3\end{bmatrix},\begin{bmatrix}-2\\3\\1\\3\\0\\-1\end{bmatrix},\begin{bmatrix}3\\-1\\1\\-7\\2\\2\end{bmatrix},\begin{bmatrix}-5\\4\\1\\2\\-4\\-5\end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix}1 & 0 & \frac{186}{131} & \frac{51}{131} & -\frac{188}{131} & \frac{77}{131}\\ 0 & 1 & -\frac{272}{131} & -\frac{45}{131} & \frac{58}{131} & -\frac{14}{131}\end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-\frac{77}{131}\\\frac{14}{131}\\0\\0\\0\\1\end{bmatrix},\begin{bmatrix}\frac{188}{131}\\-\frac{58}{131}\\0\\0\\1\\0\end{bmatrix},\begin{bmatrix}-\frac{51}{131}\\\frac{45}{131}\\0\\1\\0\\0\end{bmatrix},\begin{bmatrix}-\frac{186}{131}\\\frac{272}{131}\\1\\0\\0\\0\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\\-1\\-\frac{29}{7}\end{bmatrix},\begin{bmatrix}0\\1\\0\\0\\-\frac{11}{2}\\-\frac{94}{7}\end{bmatrix},\begin{bmatrix}0\\0\\1\\0\\10\\22\end{bmatrix},\begin{bmatrix}0\\0\\0\\1\\\frac{3}{2}\\3\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\2\\0\\5\\0\\0\\1\\-2\\3\end{bmatrix},\begin{bmatrix}0\\0\\1\\-2\\0\\0\\3\\5\\-6\end{bmatrix},\begin{bmatrix}0\\0\\0\\0\\1\\0\\1\\1\\-1\end{bmatrix},\begin{bmatrix}0\\0\\0\\0\\0\\1\\0\\-2\\-3\end{bmatrix}\right\}\right)$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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:

$$\begin{bmatrix} 10 & 18 & 24 & 24 & -12 \\ 12 & -2 & -6 & 0 & -18 \\ -30 & -21 & -23 & -30 & 39 \\ 27 & 30 & 36 & 37 & -30 \\ 18 & 24 & 30 & 30 & -20 \end{bmatrix}$$

□ Matrix brought to reduced row-echelon form:

$$\begin{bmatrix} \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & \boxed{1} & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

□ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 5 \qquad D = \{1,\,2,\,3,\,4,\,5\} \qquad F = \{\ \}$$

□ Matrix (coefficient matrix) is nonsingular or singular? (Theorem NSRRI [76]) 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$\mathcal{S}p(\{\ \})$

□ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}10\\12\\-30\\27\\18\end{bmatrix},\,\begin{bmatrix}18\\-2\\-21\\30\\24\end{bmatrix},\,\begin{bmatrix}24\\-6\\-23\\36\\30\end{bmatrix},\,\begin{bmatrix}24\\0\\-30\\37\\30\end{bmatrix},\,\begin{bmatrix}-12\\-18\\39\\-30\\-20\end{bmatrix}\right\}\right)$$

□ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$K = []$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\}\right)$$

□ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\}\right)$$

□ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\}\right)$$

□ 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 [216], Theorem NSI [228])

$$\begin{bmatrix}1 & -\left(\frac{9}{4}\right) & -\left(\frac{3}{2}\right) & 3 & -6\\ \frac{21}{2} & \frac{43}{4} & \frac{21}{2} & 9 & -9\\ -15 & -\left(\frac{21}{2}\right) & -11 & -15 & \frac{39}{2}\\ 9 & \frac{15}{4} & \frac{9}{2} & 10 & -15\\ \frac{9}{2} & \frac{3}{4} & \frac{3}{2} & 6 & -\left(\frac{19}{2}\right)\end{bmatrix}$$

□ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 [328]). (Product of all eigenvalues?)

Determinant = 16

□ Eigenvalues, and bases for eigenspaces. (Definition EEM [331],Definition EM [341])

$$\lambda = -2 \qquad E_K(-2) = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\-2\\1\\0\\1\end{bmatrix},\ \begin{bmatrix}-1\\2\\-2\\1\\0\end{bmatrix}\right\}\right)$$

$$\lambda = 1 \qquad E_K(1) = \mathcal{S}p\left(\left\{\begin{bmatrix}4\\-10\\7\\0\\2\end{bmatrix},\ \begin{bmatrix}-4\\18\\-17\\5\\0\end{bmatrix}\right\}\right)$$

$$\lambda = 4 \qquad E_K(4) = \mathcal{S}p\left(\left\{\begin{bmatrix}1\\-1\\0\\1\\1\end{bmatrix}\right\}\right)$$

☐ Geometric and algebraic multiplicities. (Definition GME [344]Definition AME [343])

$$\begin{aligned}\gamma_K(-2) &= 2 & \alpha_K(-2) &= 2\\ \gamma_K(1) &= 2 & \alpha_K(1) &= 2\\ \gamma_K(4) &= 1 & \alpha_K(4) &= 1\end{aligned}$$

☐ Diagonalizable? (Definition DZM [369])

Yes, large eigenspaces, Theorem DMLE [373].

☐ The diagonalization. (Theorem DC [370])

$$\begin{bmatrix}-4 & -3 & -4 & -6 & 7\\ -7 & -5 & -6 & -8 & 10\\ 1 & -1 & -1 & 1 & -3\\ 1 & 0 & 0 & 1 & -2\\ 2 & 5 & 6 & 4 & 0\end{bmatrix}\begin{bmatrix}10 & 18 & 24 & 24 & -12\\ 12 & -2 & -6 & 0 & -18\\ -30 & -21 & -23 & -30 & 39\\ 27 & 30 & 36 & 37 & -30\\ 18 & 24 & 30 & 30 & -20\end{bmatrix}\begin{bmatrix}2 & -1 & 4 & -4 & 1\\ -2 & 2 & -10 & 18 & -1\\ 1 & -2 & 7 & -17 & 0\\ 0 & 1 & 0 & 5 & 1\\ 1 & 0 & 2 & 0 & 1\end{bmatrix}$$

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

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:

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

☐ Analysis of the row-reduced matrix (Notation RREFA [51]):

$$r = 5 \qquad D = \{1,\,2,\,3\} \qquad F = \{4,\,5\}$$

☐ Matrix (coefficient matrix) is nonsingular or singular? (Theorem NSRRI [76]) 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 [121], Theorem BNS [138]). Solve the homogenous system with this matrix as the coefficient matrix and write the solutions in vector form (Theorem VFSLS [104]) to see these vectors arise.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-1\\2\\-2\\1\\0\end{bmatrix},\,\begin{bmatrix}2\\-2\\1\\0\\1\end{bmatrix}\right\}\right)$$

☐ Range 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 BROC [172])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}- - 2\\-6\\10\\-7\\-4\end{bmatrix},\,\begin{bmatrix}-1\\-5\\7\\-5\\-3\end{bmatrix},\,\begin{bmatrix}-2\\-4\\7\\-6\\-4\end{bmatrix}\right\}\right)$$

☐ The range of the matrix, as it arises from row operations on an augmented matrix. The matrix K is computed as described in Theorem RNS [177]. This is followed by the range described by a set of linearly independent vectors that span the null space of K, computed as according to Theorem SSNS [121]. When $r = m$, the matrix K has no rows and the range is all of $\mathbb{C}^m$.

$$K = \begin{bmatrix}1 & 0 & -2 & -6 & 5\\0 & 1 & 4 & 10 & -9\end{bmatrix}$$

$$\mathcal{S}p\left(\left\{\begin{bmatrix}-5\\9\\0\\0\\1\end{bmatrix},\,\begin{bmatrix}6\\-10\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}2\\-4\\1\\0\\0\end{bmatrix}\right\}\right)$$

☐ Range of the matrix, expressed as the span of a set of linearly independent vectors. These vectors are computed by row-reducing the transpose the matrix into reduced row-echelon form, tossing out the zero rows, and writing the remaining rows as column vectors. By Theorem RMRST [188] and Theorem BRS [186], and in the style of Example IS [24], this yields a linearly independent set of vectors that span the range.

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\\frac{9}{4}\\\frac{5}{2}\end{bmatrix},\ \begin{bmatrix}0\\1\\0\\\frac{5}{4}\\\frac{3}{2}\end{bmatrix},\ \begin{bmatrix}0\\0\\1\\\frac{1}{2}\\1\end{bmatrix}\right\}\right)$$

☐ 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 [186])

$$\mathcal{S}p\left(\left\{\begin{bmatrix}1\\0\\0\\1\\-2\end{bmatrix},\ \begin{bmatrix}0\\1\\0\\-2\\2\end{bmatrix},\ \begin{bmatrix}0\\0\\1\\2\\-1\end{bmatrix}\right\}\right)$$

☐ 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 [216], Theorem NSI [228])

☐ Subspace dimensions associated with the matrix. (Definition NOM [300], Definition ROM [300]) Verify Theorem RPNC [302]

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 [328]). (Product of all eigenvalues?)

Determinant $=$ 0

☐ Eigenvalues, and bases for eigenspaces. (Definition EEM [331],Definition EM [341])

$$\lambda = -1 \qquad E_L(-1) = \mathcal{S}p\left(\left\{\begin{bmatrix}-5\\9\\0\\0\\1\end{bmatrix}, \begin{bmatrix}6\\-10\\0\\1\\0\end{bmatrix}, \begin{bmatrix}2\\-4\\1\\0\\0\end{bmatrix}\right\}\right)$$

$$\lambda = 0 \qquad E_L(0) = \mathcal{S}p\left(\left\{\begin{bmatrix}2\\-2\\1\\0\\1\end{bmatrix}, \begin{bmatrix}-1\\2\\-2\\1\\0\end{bmatrix}\right\}\right)$$

□ Geometric and algebraic multiplicities. (Definition GME [344]Definition AME [343])

$$\gamma_L(-1) = 3 \qquad \alpha_L(-1) = 3$$
$$\gamma_L(0) = 2 \qquad \alpha_L(0) = 2$$

□ Diagonalizable? (Definition DZM [369])

Yes, large eigenspaces, Theorem DMLE [373].

□ The diagonalization. (Theorem DC [370])

$$\begin{bmatrix}4 & 3 & 4 & 6 & -6\\7 & 5 & 6 & 9 & -10\\-10 & -7 & -7 & -10 & 13\\-4 & -3 & -4 & -6 & 7\\-7 & -5 & -6 & -8 & 10\end{bmatrix}\begin{bmatrix}-2 & -1 & -2 & -4 & 4\\-6 & -5 & -4 & -4 & 6\\10 & 7 & 7 & 10 & -13\\-7 & -5 & -6 & -9 & 10\\-4 & -3 & -4 & -6 & 6\end{bmatrix}\begin{bmatrix}-5 & 6 & 2 & 2 & -1\\9 & -10 & -4 & -2 & 2\\0 & 0 & 1 & 1 & -2\\0 & 1 & 0 & 0 & 1\\1 & 0 & 0 & 1 & 0\end{bmatrix}$$
$$= \begin{bmatrix}-1 & 0 & 0 & 0 & 0\\0 & -1 & 0 & 0 & 0\\0 & 0 & -1 & 0 & 0\\0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\end{bmatrix}$$

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 [379])

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^3, \quad T\left(\begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix}\right) = \begin{bmatrix}x_1+2x_2+3x_3+4x_4+4x_5\\3x_1+x_2+4x_3-3x_4+7x_5\\x_1-x_2-5x_4+x_5\end{bmatrix}$$

□ A basis for the null space of the linear transformation: (Definition NSLT [403])

$$\left\{\begin{bmatrix}-2\\-1\\0\\0\\1\end{bmatrix},\ \begin{bmatrix}2\\-3\\0\\1\\0\end{bmatrix},\ \begin{bmatrix}-1\\-1\\1\\0\\0\end{bmatrix}\right\}$$

□ Injective: No. (Definition ILT [399])

Since the null space is nontrivial Theorem XXoneonetrivial [**??**] 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

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

and

$$\mathbf{z} = \begin{bmatrix} -1 \\ -5 \\ 1 \\ 1 \\ 1 \end{bmatrix} \in N(T)$$

so the vector $\mathbf{z}$ effectively "does nothing" in the evaluation of T.

□ A basis for the range of the linear transformation: (Definition RLT [416])

Evaluate the linear transformation on a standard basis to get a spanning set for the range:

$$\left\{ \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \\ -5 \end{bmatrix}, \begin{bmatrix} 4 \\ 7 \\ 1 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (XXXX). If this spanning set is not linearly dependent at first, use relations of linear dependence on the set to remove redundant elements while preserving the spanning properties. Continue this until the set is linear independent. 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 [411])

Notice that the range is not all of $\mathbb{C}^3$ since its 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. 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.

Domain dimension: 5 Rank: 2 Nullity: 3

□ Invertible: No.

Not injective or surjective.

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 [379])

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^3, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + x_2 + 3x_3 - 4x_4 + 5x_5 \\ x_1 - 2x_2 + 3x_3 - 9x_4 + 3x_5 \\ 3x_1 + 4x_3 - 6x_4 + 5x_5 \end{bmatrix}$$

□ A basis for the null space of the linear transformation: (Definition NSLT [403])

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 3 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Since the null space is nontrivial Theorem XXoneonetrivial [**??**] 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

$$\begin{bmatrix} -4 \\ -4 \\ -2 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \\ -2 \\ -3 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -5 \\ 0 \\ 2 \\ 3 \end{bmatrix}$$

and

$$\mathbf{z} = \begin{bmatrix} -1 \\ -5 \\ 0 \\ 2 \\ 3 \end{bmatrix} \in N(T)$$

so the vector $\mathbf{z}$ effectively "does nothing" in the evaluation of T. □ A basis for the range of the linear transformation: (Definition RLT [416])

Evaluate the linear transformation on a standard basis to get a spanning set for the range:

$$\left\{ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -4 \\ -9 \\ -6 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 5 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (XXXX). If this spanning set is not linearly dependent at first, use relations of linear dependence on the set to remove redundant elements while preserving the spanning properties. Continue this until the set is linear independent. 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 [411])

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.

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.

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 [379])

$$T\colon \mathbb{C}^3 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}\right) = \begin{bmatrix}-x_1+x_2-3x_3\\-x_1+2x_2-4x_3\\x_1+x_2+x_3\\2x_1+3x_2+x_3\\x_1+2x_3\end{bmatrix}$$

□ A basis for the null space of the linear transformation: (Definition NSLT [403])

$$\left\{\begin{bmatrix}-2\\1\\1\end{bmatrix}\right\}$$

□ Injective: No. (Definition ILT [399])

Since the null space is nontrivial Theorem XXoneonetrivial [**??**] 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

$$\mathbf{z} = \begin{bmatrix}-4\\2\\2\end{bmatrix} \in N(T)$$

so the vector $\mathbf{z}$ effectively "does nothing" in the evaluation of T.

☐ A basis for the range of the linear transformation: (Definition RLT [416])

Evaluate the linear transformation on a standard basis to get a spanning set for the range:

$$\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ 1 \\ 1 \\ 2 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (XXXX). If this spanning set is not linearly dependent at first, use relations of linear dependence on the set to remove redundant elements while preserving the spanning properties. Continue this until the set is linear independent. 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.

Domain dimension: 3 Rank: 2 Nullity: 1

☐ Surjective: No. (Definition SLT [411])

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 $\begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \not\in \mathcal{R}(T)$, by setting the output equal to this vector and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent. This alone is sufficient to see that the linear transformation is not onto.

☐ Invertible: No.

Not injective, and the relative dimensions of the domain and codomain prohibit any possibility of being surjective.

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 [379])

$$T\colon \mathbb{C}^3 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -x_1 + x_2 + x_3 \\ -x_1 + 2x_2 + 2x_3 \\ x_1 + x_2 + 3x_3 \\ 2x_1 + 3x_2 + x_3 \\ -2x_1 + x_2 + 3x_3 \end{bmatrix}$$

□ A basis for the null space of the linear transformation: (Definition NSLT [403])

$$\{\,\}$$

□ Injective: Yes. (Definition ILT [399])

Since $\mathcal{N}(T) = \{\mathbf{0}\}$, Theorem XXoneonetrivial [**??**] tells us that T is injective.

□ A basis for the range of the linear transformation: (Definition RLT [416])

Evaluate the linear transformation on a standard basis to get a spanning set for the range:

$$\left\{ \begin{bmatrix} -1 \\ -1 \\ 1 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 1 \\ 3 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (XXXX). If this spanning set is not linearly dependent at first, use relations of linear dependence on the set to remove redundant elements while preserving the spanning properties. Continue this until the set is linear independent. 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 [411])

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. 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.

Domain dimension: 3 Rank: 3 Nullity: 0

☐ Invertible: No.

Not surjective.

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 [379])

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} -2x_1 + 3x_2 + 3x_3 - 6x_4 + 3x_5 \\ -16x_1 + 9x_2 + 12x_3 - 28x_4 + 28x_5 \\ -19x_1 + 7x_2 + 14x_3 - 32x_4 + 37x_5 \\ -21x_1 + 9x_2 + 15x_3 - 35x_4 + 39x_5 \\ -9x_1 + 5x_2 + 7x_3 - 16x_4 + 16x_5 \end{bmatrix}$$

□ A basis for the null space of the linear transformation: (Definition NSLT [403])

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

□ Injective: No. (Definition ILT [399])

Since the null space is nontrivial Theorem XXoneonetrivial [**??**] 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

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

and

$$\mathbf{z} = \begin{bmatrix} 3 \\ 4 \\ 1 \\ 3 \\ 3 \end{bmatrix} \in N(T)$$

so the vector $\mathbf{z}$ effectively "does nothing" in the evaluation of T.

□ A basis for the range of the linear transformation: (Definition RLT [416])

Evaluate the linear transformation on a standard basis to get a spanning set for the range:

$$\left\{ \begin{bmatrix} -2 \\ -16 \\ -19 \\ -21 \\ -9 \end{bmatrix}, \begin{bmatrix} 3 \\ 9 \\ 7 \\ 9 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 12 \\ 14 \\ 15 \\ 7 \end{bmatrix}, \begin{bmatrix} -6 \\ -28 \\ -32 \\ -35 \\ -16 \end{bmatrix}, \begin{bmatrix} 3 \\ 28 \\ 37 \\ 39 \\ 16 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (XXXX). If this spanning set is not linearly dependent at first, use relations of linear dependence on the set to remove redundant elements while preserving the spanning properties. Continue this until the set is linear independent. A basis for the range is:

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right\}$$

□ Surjective: No. (Definition SLT [411])

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 XXrangecodomain [**??**] the transformation is not surjective.

To be more precise, verify that $\begin{bmatrix} -1 \\ 2 \\ 3 \\ -1 \\ 4 \end{bmatrix} \notin R(T)$, by setting the output equal to this vector and seeing that the resulting system of linear equations has no solution, i.e. is inconsistent. 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.

Domain dimension: 5 Rank: 4 Nullity: 1

□ Invertible: No.

Neither injective nor surjective. Notice that since the domain and codomain have the same dimesion, 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 XXsumof rankandnullity [**??**].

Archetype R

Summary Linear transformation with equal-sized domain and codomain. Injective, surjective, invertible, diagonalizable, the works.

□ A linear transformation: (Definition LT [379])

$$T\colon \mathbb{C}^5 \mapsto \mathbb{C}^5, \quad T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}\right) = \begin{bmatrix} -65x_1 + 128x_2 + 10x_3 - 262x_4 + 40x_5 \\ 36x_1 - 73x_2 - x_3 + 151x_4 - 16x_5 \\ -44x_1 + 88x_2 + 5x_3 - 180x_4 + 24x_5 \\ 34x_1 - 68x_2 - 3x_3 + 140x_4 - 18x_5 \\ 12x_1 - 24x_2 - x_3 + 49x_4 - 5x_5 \end{bmatrix}$$

□ A basis for the null space of the linear transformation: (Definition NSLT [403])

$$\{\ \}$$

□ Injective: Yes. (Definition ILT [399])

Since the null space is trivial Theorem XXoneonetrivial [**??**] tells us that the linear transformation is injective.

□ A basis for the range of the linear transformation: (Definition RLT [416])

Evaluate the linear transformation on a standard basis to get a spanning set for the range:

$$\left\{ \begin{bmatrix} -65 \\ 36 \\ -44 \\ 34 \\ 12 \end{bmatrix}, \begin{bmatrix} 128 \\ -73 \\ 88 \\ -68 \\ -24 \end{bmatrix}, \begin{bmatrix} 10 \\ -1 \\ 5 \\ -3 \\ -1 \end{bmatrix}, \begin{bmatrix} -262 \\ 151 \\ -180 \\ 140 \\ 49 \end{bmatrix}, \begin{bmatrix} 40 \\ -16 \\ 24 \\ -18 \\ -5 \end{bmatrix} \right\}$$

If the linear transformation is injective, then the set above is guaranteed to be linearly independent (XXXX). If this spanning set is not linearly dependent at first, use relations of linear dependence on the set to remove redundant elements while preserving the spanning properties. Continue this until the set is linear independent. A basis for the range is:

$$\left\{\begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix},\,\begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix}\right\}$$

□ Surjective: Yes/No. (Definition SLT [411])

A basis for the range is the standard basis of $\mathbb{C}^5$, so $\mathcal{R}(T) = \mathbb{C}^5$ and Theorem XXrange-codomain [**??**] 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.

Domain dimension: 5 Rank: 5 Nullity: 0

□ Invertible: Yes.

Both injective and surjective. Notice that since the domain and codomain have the same dimesion, 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.

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 [379])

$$T\colon \mathbb{C}^3 \mapsto M_{22}, \quad T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = \begin{bmatrix} a-b & 2a+2b+c \\ 3a+b+c & -2a-6b-2c \end{bmatrix}$$

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 [379])

$$T\colon P_4 \mapsto P_5, \quad T\left(p(x)\right) = (x-2)p(x)$$

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 [379])

$$T\colon M_{23} \mapsto \mathbb{C}^4, \quad T\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a+2b+12c-3d+e+6f \\ 2a-b-c+d-11f \\ a+b+7c+2d+e-3f \\ a+2b+12c+5e-5f \end{bmatrix}$$

Archetype V

Summary Domain is polynomials, codomain is matrices. Domain and codomain both have dimension 4. Injective, surjective, invertible, (eigenvalues, diagonalizable???).

□ A linear transformation: (Definition LT [379])

$$T\colon P_3 \mapsto M_{22}, \quad T\left(a + bx + cx^2 + dx^3\right) = \begin{bmatrix} a+b & a-2c \\ d & b-d \end{bmatrix}$$

□ When invertible, the inverse linear transformation. (Definition IVLT [425])

$$T^{-1}\colon M_{22} \mapsto P_3, \quad T^{-1}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a-c-d) + (c+d)x + \frac{1}{2}(a-b-c-d)x^2 + cx^3$$

Part T
Topics

P: Preliminaries

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

$$
\begin{aligned}
(2+5i)+(6-4i) &= (2+6)+(5+(-4))i = 8+i \\
(2+5i)-(6-4i) &= (2-6)+(5-(-4))i = -4+9i \\
(2+5i)(6-4i) &= (2)(6)+(5i)(6)+(2)(-4i)+(5i)(-4i) = 12+30i-8i-20i^2 \\
&= 12+22i-20(-1) = 32+22i
\end{aligned}
$$

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.

$$\frac{2+5i}{6-4i} = \frac{2+5i}{6-4i}\frac{6+4i}{6+4i} = \frac{-8+38i}{52} = -\frac{8}{52}+\frac{38}{52}i = -\frac{2}{13}+\frac{19}{26}i \qquad \circledcirc$$

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 now define.

Subsection CCN
Conjugates of Complex Numbers

Definition CCN
Conjugate of a Complex Number
The **conjugate** of the complex number $c = a+bi \in \mathbb{C}$ is the complex number $\overline{c} = a-bi$. △

Example CSCN
Conjugate of some complex numbers

$$\overline{2+3i} = 2-3i \qquad \overline{5-4i} = 5+4i \qquad \overline{-3+0i} = -3+0i \qquad \overline{0+0i} = 0+0i$$ ⊚

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 c and d are complex numbers. Then $\overline{c+d} = \overline{c} + \overline{d}$. □

Proof Let $c = a+bi$ and $d = r+si$. Then

$$\overline{c+d} = \overline{(a+r)+(b+s)i} = (a+r)-(b+s)i = (a-bi)+(r-si) = \overline{c}+\overline{d}$$ ■

Theorem CCRM
Complex Conjugation Respects Multiplication
Suppose that c and d are complex numbers. Then $\overline{cd} = \overline{c}\,\overline{d}$. □

Proof Let $c = a+bi$ and $d = r+si$. Then

$$\begin{aligned}\overline{cd} &= \overline{(ar-bs)+(as+br)i} = (ar-bs)-(as+br)i \\ &= (ar-(-b)(-s))+(a(-s)+(-b)r)i = (a-bi)(r-si) = \overline{c}\,\overline{d}\end{aligned}$$ ■

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 $c = a+bi \in \mathbb{C}$, is the real number

$$|c| = \sqrt{c\overline{c}} = \sqrt{a^2+b^2}.$$ △

Example MSCN
Modulus of some complex numbers

$$|2+3i| = \sqrt{13} \qquad |5-4i| = \sqrt{41} \qquad |-3+0i| = 3 \qquad |0+0i| = 0 \qquad \circledcirc$$

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.

Part A
Applications

Index