Undergraduate Texts in Mathematics

UTM

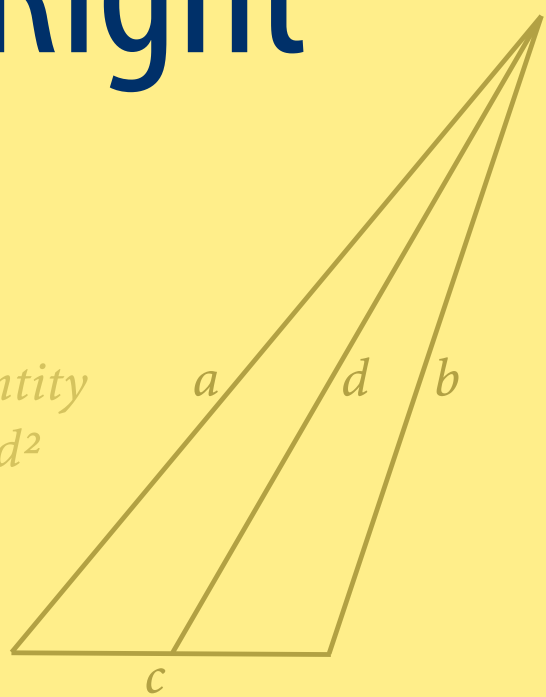
Sheldon Axler

Linear Algebra Done Right

Third Edition

Apollonius's Identity

$a^2 + b^2 = \frac{1}{2}c^2 + 2d^2$



Springer

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics

Undergraduate Texts in Mathematics are generally aimed at third- and fourth-year undergraduate mathematics students at North American universities. These texts strive to provide students and teachers with new perspectives and novel approaches. The books include motivation that guides the reader to an appreciation of interrelations among different aspects of the subject. They feature examples that illustrate key concepts as well as exercises that strengthen understanding.

For further volumes:
http://www.springer.com/series/666

Sheldon Axler

Linear Algebra Done Right

Third edition

Springer

Sheldon Axler
Department of Mathematics
San Francisco State University
San Francisco, CA, USA

ISSN 0172-6056 ISSN 2197-5604 (electronic)
ISBN 978-3-319-11079-0 ISBN 978-3-319-11080-6 (eBook)
DOI 10.1007/978-3-319-11080-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014954079

Mathematics Subject Classification (2010): 15-01, 15A03, 15A04, 15A15, 15A18, 15A21

Typeset by the author in LaTeX

Cover figure: For a statement of Apollonius's Identity and its connection to linear algebra, see the last exercise in Section 6.A.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Preface for the Instructor

You are about to teach a course that will probably give students their second exposure to linear algebra. During their first brush with the subject, your students probably worked with Euclidean spaces and matrices. In contrast, this course will emphasize abstract vector spaces and linear maps.

The audacious title of this book deserves an explanation. Almost all linear algebra books use determinants to prove that every linear operator on a finite-dimensional complex vector space has an eigenvalue. Determinants are difficult, nonintuitive, and often defined without motivation. To prove the theorem about existence of eigenvalues on complex vector spaces, most books must define determinants, prove that a linear map is not invertible if and only if its determinant equals 0, and then define the characteristic polynomial. This tortuous (torturous?) path gives students little feeling for why eigenvalues exist.

In contrast, the simple determinant-free proofs presented here (for example, see 5.21) offer more insight. Once determinants have been banished to the end of the book, a new route opens to the main goal of linear algebra—understanding the structure of linear operators.

This book starts at the beginning of the subject, with no prerequisites other than the usual demand for suitable mathematical maturity. Even if your students have already seen some of the material in the first few chapters, they may be unaccustomed to working exercises of the type presented here, most of which require an understanding of proofs.

Here is a chapter-by-chapter summary of the highlights of the book:

- Chapter 1: Vector spaces are defined in this chapter, and their basic properties are developed.

- Chapter 2: Linear independence, span, basis, and dimension are defined in this chapter, which presents the basic theory of finite-dimensional vector spaces.

- Chapter 3: Linear maps are introduced in this chapter. The key result here is the Fundamental Theorem of Linear Maps (3.22): if T is a linear map on V, then $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$. Quotient spaces and duality are topics in this chapter at a higher level of abstraction than other parts of the book; these topics can be skipped without running into problems elsewhere in the book.

- Chapter 4: The part of the theory of polynomials that will be needed to understand linear operators is presented in this chapter. This chapter contains no linear algebra. It can be covered quickly, especially if your students are already familiar with these results.

- Chapter 5: The idea of studying a linear operator by restricting it to small subspaces leads to eigenvectors in the early part of this chapter. The highlight of this chapter is a simple proof that on complex vector spaces, eigenvalues always exist. This result is then used to show that each linear operator on a complex vector space has an upper-triangular matrix with respect to some basis. All this is done without defining determinants or characteristic polynomials!

- Chapter 6: Inner product spaces are defined in this chapter, and their basic properties are developed along with standard tools such as orthonormal bases and the Gram–Schmidt Procedure. This chapter also shows how orthogonal projections can be used to solve certain minimization problems.

- Chapter 7: The Spectral Theorem, which characterizes the linear operators for which there exists an orthonormal basis consisting of eigenvectors, is the highlight of this chapter. The work in earlier chapters pays off here with especially simple proofs. This chapter also deals with positive operators, isometries, the Polar Decomposition, and the Singular Value Decomposition.

- Chapter 8: Minimal polynomials, characteristic polynomials, and generalized eigenvectors are introduced in this chapter. The main achievement of this chapter is the description of a linear operator on a complex vector space in terms of its generalized eigenvectors. This description enables one to prove many of the results usually proved using Jordan Form. For example, these tools are used to prove that every invertible linear operator on a complex vector space has a square root. The chapter concludes with a proof that every linear operator on a complex vector space can be put into Jordan Form.

- Chapter 9: Linear operators on real vector spaces occupy center stage in this chapter. Here the main technique is complexification, which is a natural extension of an operator on a real vector space to an operator on a complex vector space. Complexification allows our results about complex vector spaces to be transferred easily to real vector spaces. For example, this technique is used to show that every linear operator on a real vector space has an invariant subspace of dimension 1 or 2. As another example, we show that that every linear operator on an odd-dimensional real vector space has an eigenvalue.

- Chapter 10: The trace and determinant (on complex vector spaces) are defined in this chapter as the sum of the eigenvalues and the product of the eigenvalues, both counting multiplicity. These easy-to-remember definitions would not be possible with the traditional approach to eigenvalues, because the traditional method uses determinants to prove that sufficient eigenvalues exist. The standard theorems about determinants now become much clearer. The Polar Decomposition and the Real Spectral Theorem are used to derive the change of variables formula for multivariable integrals in a fashion that makes the appearance of the determinant there seem natural.

This book usually develops linear algebra simultaneously for real and complex vector spaces by letting **F** denote either the real or the complex numbers. If you and your students prefer to think of **F** as an arbitrary field, then see the comments at the end of Section 1.A. I prefer avoiding arbitrary fields at this level because they introduce extra abstraction without leading to any new linear algebra. Also, students are more comfortable thinking of polynomials as functions instead of the more formal objects needed for polynomials with coefficients in finite fields. Finally, even if the beginning part of the theory were developed with arbitrary fields, inner product spaces would push consideration back to just real and complex vector spaces.

You probably cannot cover everything in this book in one semester. Going through the first eight chapters is a good goal for a one-semester course. If you must reach Chapter 10, then consider covering Chapters 4 and 9 in fifteen minutes each, as well as skipping the material on quotient spaces and duality in Chapter 3.

A goal more important than teaching any particular theorem is to develop in students the ability to understand and manipulate the objects of linear algebra. Mathematics can be learned only by doing. Fortunately, linear algebra has many good homework exercises. When teaching this course, during each class I usually assign as homework several of the exercises, due the next class. Going over the homework might take up a third or even half of a typical class.

Major changes from the previous edition:

- This edition contains 561 exercises, including 337 new exercises that were not in the previous edition. Exercises now appear at the end of each section, rather than at the end of each chapter.
- Many new examples have been added to illustrate the key ideas of linear algebra.
- Beautiful new formatting, including the use of color, creates pages with an unusually pleasant appearance in both print and electronic versions. As a visual aid, definitions are in beige boxes and theorems are in blue boxes (in color versions of the book).
- Each theorem now has a descriptive name.
- New topics covered in the book include product spaces, quotient spaces, and duality.
- Chapter 9 (Operators on Real Vector Spaces) has been completely rewritten to take advantage of simplifications via complexification. This approach allows for more streamlined presentations in Chapters 5 and 7 because those chapters now focus mostly on complex vector spaces.
- Hundreds of improvements have been made throughout the book. For example, the proof of Jordan Form (Section 8.D) has been simplified.

Please check the website below for additional information about the book. I may occasionally write new sections on additional topics. These new sections will be posted on the website. Your suggestions, comments, and corrections are most welcome.

Best wishes for teaching a successful linear algebra class!

Sheldon Axler
Mathematics Department
San Francisco State University
San Francisco, CA 94132, USA

website: `linear.axler.net`
e-mail: `linear@axler.net`
Twitter: `@AxlerLinear`

Preface for the Student

You are probably about to begin your second exposure to linear algebra. Unlike your first brush with the subject, which probably emphasized Euclidean spaces and matrices, this encounter will focus on abstract vector spaces and linear maps. These terms will be defined later, so don't worry if you do not know what they mean. This book starts from the beginning of the subject, assuming no knowledge of linear algebra. The key point is that you are about to immerse yourself in serious mathematics, with an emphasis on attaining a deep understanding of the definitions, theorems, and proofs.

You cannot read mathematics the way you read a novel. If you zip through a page in less than an hour, you are probably going too fast. When you encounter the phrase "as you should verify", you should indeed do the verification, which will usually require some writing on your part. When steps are left out, you need to supply the missing pieces. You should ponder and internalize each definition. For each theorem, you should seek examples to show why each hypothesis is necessary. Discussions with other students should help.

As a visual aid, definitions are in beige boxes and theorems are in blue boxes (in color versions of the book). Each theorem has a descriptive name.

Please check the website below for additional information about the book. I may occasionally write new sections on additional topics. These new sections will be posted on the website. Your suggestions, comments, and corrections are most welcome.

Best wishes for success and enjoyment in learning linear algebra!

Sheldon Axler
Mathematics Department
San Francisco State University
San Francisco, CA 94132, USA

website: `linear.axler.net`
e-mail: `linear@axler.net`
Twitter: `@AxlerLinear`

Acknowledgments

I owe a huge intellectual debt to the many mathematicians who created linear algebra over the past two centuries. The results in this book belong to the common heritage of mathematics. A special case of a theorem may first have been proved in the nineteenth century, then slowly sharpened and improved by many mathematicians. Bestowing proper credit on all the contributors would be a difficult task that I have not undertaken. In no case should the reader assume that any theorem presented here represents my original contribution. However, in writing this book I tried to think about the best way to present linear algebra and to prove its theorems, without regard to the standard methods and proofs used in most textbooks.

Many people helped make this a better book. The two previous editions of this book were used as a textbook at about 300 universities and colleges. I received thousands of suggestions and comments from faculty and students who used the second edition. I looked carefully at all those suggestions as I was working on this edition. At first, I tried keeping track of whose suggestions I used so that those people could be thanked here. But as changes were made and then replaced with better suggestions, and as the list grew longer, keeping track of the sources of each suggestion became too complicated. And lists are boring to read anyway. Thus in lieu of a long list of people who contributed good ideas, I will just say how truly grateful I am to everyone who sent me suggestions and comments. Many many thanks!

Special thanks to Ken Ribet and his giant (220 students) linear algebra class at Berkeley that class-tested a preliminary version of this third edition and that sent me more suggestions and corrections than any other group.

Finally, I thank Springer for providing me with help when I needed it and for allowing me the freedom to make the final decisions about the content and appearance of this book. Special thanks to Elizabeth Loew for her wonderful work as editor and David Kramer for unusually skillful copyediting.

Sheldon Axler

CHAPTER

1

René Descartes explaining his work to Queen Christina of Sweden. Vector spaces are a generalization of the description of a plane using two coordinates, as published by Descartes in 1637.

Vector Spaces

Linear algebra is the study of linear maps on finite-dimensional vector spaces. Eventually we will learn what all these terms mean. In this chapter we will define vector spaces and discuss their elementary properties.

In linear algebra, better theorems and more insight emerge if complex numbers are investigated along with real numbers. Thus we will begin by introducing the complex numbers and their basic properties.

We will generalize the examples of a plane and ordinary space to $\mathbf{R}^n$ and $\mathbf{C}^n$, which we then will generalize to the notion of a vector space. The elementary properties of a vector space will already seem familiar to you.

Then our next topic will be subspaces, which play a role for vector spaces analogous to the role played by subsets for sets. Finally, we will look at sums of subspaces (analogous to unions of subsets) and direct sums of subspaces (analogous to unions of disjoint sets).

LEARNING OBJECTIVES FOR THIS CHAPTER

- basic properties of the complex numbers
- $\mathbf{R}^n$ and $\mathbf{C}^n$
- vector spaces
- subspaces
- sums and direct sums of subspaces

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_1

1.A $\mathbf{R}^n$ and $\mathbf{C}^n$

Complex Numbers

You should already be familiar with basic properties of the set $\mathbf{R}$ of real numbers. Complex numbers were invented so that we can take square roots of negative numbers. The idea is to assume we have a square root of -1, denoted i, that obeys the usual rules of arithmetic. Here are the formal definitions:

1.1 Definition ***complex numbers***

- A ***complex number*** is an ordered pair (a, b), where $a, b \in \mathbf{R}$, but we will write this as $a + bi$.
- The set of all complex numbers is denoted by $\mathbf{C}$:
$$\mathbf{C} = \{a + bi : a, b \in \mathbf{R}\}.$$
- ***Addition and multiplication*** on $\mathbf{C}$ are defined by
$$(a + bi) + (c + di) = (a + c) + (b + d)i,$$
$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i;$$
here $a, b, c, d \in \mathbf{R}$.

If $a \in \mathbf{R}$, we identify $a + 0i$ with the real number a. Thus we can think of $\mathbf{R}$ as a subset of $\mathbf{C}$. We also usually write $0 + bi$ as just bi, and we usually write $0 + 1i$ as just i.

The symbol i was first used to denote $\sqrt{-1}$ by Swiss mathematician Leonhard Euler in 1777.

Using multiplication as defined above, you should verify that $i^2 = -1$. Do not memorize the formula for the product of two complex numbers; you can always rederive it by recalling that $i^2 = -1$ and then using the usual rules of arithmetic (as given by 1.3).

1.2 Example Evaluate $(2 + 3i)(4 + 5i)$.

Solution
$$\begin{aligned}(2 + 3i)(4 + 5i) &= 2 \cdot 4 + 2 \cdot (5i) + (3i) \cdot 4 + (3i)(5i) \\ &= 8 + 10i + 12i - 15 \\ &= -7 + 22i\end{aligned}$$

1.3 Properties of complex arithmetic

commutativity
$\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbf{C}$;

associativity
$(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ and $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$;

identities
$\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$ for all $\lambda \in \mathbf{C}$;

additive inverse
for every $\alpha \in \mathbf{C}$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha + \beta = 0$;

multiplicative inverse
for every $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha\beta = 1$;

distributive property
$\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$.

The properties above are proved using the familiar properties of real numbers and the definitions of complex addition and multiplication. The next example shows how commutativity of complex multiplication is proved. Proofs of the other properties above are left as exercises.

1.4 **Example** Show that $\alpha\beta = \beta\alpha$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

Solution Suppose $\alpha = a + bi$ and $\beta = c + di$, where $a, b, c, d \in \mathbf{R}$. Then the definition of multiplication of complex numbers shows that

$$\begin{aligned} \alpha\beta &= (a + bi)(c + di) \\ &= (ac - bd) + (ad + bc)i \end{aligned}$$

and

$$\begin{aligned} \beta\alpha &= (c + di)(a + bi) \\ &= (ca - db) + (cb + da)i. \end{aligned}$$

The equations above and the commutativity of multiplication and addition of real numbers show that $\alpha\beta = \beta\alpha$.

1.5 Definition ***$-\alpha$, subtraction, $1/\alpha$, division***

Let $\alpha, \beta \in \mathbf{C}$.

- Let $-\alpha$ denote the additive inverse of α. Thus $-\alpha$ is the unique complex number such that
$$\alpha + (-\alpha) = 0.$$
- ***Subtraction*** on $\mathbf{C}$ is defined by
$$\beta - \alpha = \beta + (-\alpha).$$
- For $\alpha \neq 0$, let $1/\alpha$ denote the multiplicative inverse of α. Thus $1/\alpha$ is the unique complex number such that
$$\alpha(1/\alpha) = 1.$$
- ***Division*** on $\mathbf{C}$ is defined by
$$\beta/\alpha = \beta(1/\alpha).$$

So that we can conveniently make definitions and prove theorems that apply to both real and complex numbers, we adopt the following notation:

1.6 Notation $\mathbf{F}$

Throughout this book, $\mathbf{F}$ stands for either $\mathbf{R}$ or $\mathbf{C}$.

The letter $\mathbf{F}$ is used because $\mathbf{R}$ and $\mathbf{C}$ are examples of what are called fields.

Thus if we prove a theorem involving $\mathbf{F}$, we will know that it holds when $\mathbf{F}$ is replaced with $\mathbf{R}$ and when $\mathbf{F}$ is replaced with $\mathbf{C}$.

Elements of $\mathbf{F}$ are called ***scalars***. The word "scalar", a fancy word for "number", is often used when we want to emphasize that an object is a number, as opposed to a vector (vectors will be defined soon).

For $\alpha \in \mathbf{F}$ and m a positive integer, we define α^m to denote the product of α with itself m times:
$$\alpha^m = \underbrace{\alpha \cdots \alpha}_{m \text{ times}}.$$
Clearly $(\alpha^m)^n = \alpha^{mn}$ and $(\alpha\beta)^m = \alpha^m \beta^m$ for all $\alpha, \beta \in \mathbf{F}$ and all positive integers m, n.

Lists

Before defining $\mathbf{R}^n$ and $\mathbf{C}^n$, we look at two important examples.

1.7 Example $\mathbf{R}^2$ *and* $\mathbf{R}^3$

- The set $\mathbf{R}^2$, which you can think of as a plane, is the set of all ordered pairs of real numbers:
$$\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}.$$
- The set $\mathbf{R}^3$, which you can think of as ordinary space, is the set of all ordered triples of real numbers:
$$\mathbf{R}^3 = \{(x, y, z) : x, y, z \in \mathbf{R}\}.$$

To generalize $\mathbf{R}^2$ and $\mathbf{R}^3$ to higher dimensions, we first need to discuss the concept of lists.

1.8 Definition ***list, length***

Suppose n is a nonnegative integer. A ***list*** of ***length*** n is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses. A list of length n looks like this:
$$(x_1, \ldots, x_n).$$
Two lists are equal if and only if they have the same length and the same elements in the same order.

Thus a list of length 2 is an ordered pair, and a list of length 3 is an ordered triple.

Many mathematicians call a list of length n an ***n-tuple***.

Sometimes we will use the word ***list*** without specifying its length. Remember, however, that by definition each list has a finite length that is a nonnegative integer. Thus an object that looks like
$$(x_1, x_2, \ldots),$$
which might be said to have infinite length, is not a list.

A list of length 0 looks like this: $(\,)$. We consider such an object to be a list so that some of our theorems will not have trivial exceptions.

Lists differ from sets in two ways: in lists, order matters and repetitions have meaning; in sets, order and repetitions are irrelevant.

1.9 Example ***lists versus sets***

- The lists $(3,5)$ and $(5,3)$ are not equal, but the sets $\{3,5\}$ and $\{5,3\}$ are equal.
- The lists $(4,4)$ and $(4,4,4)$ are not equal (they do not have the same length), although the sets $\{4,4\}$ and $\{4,4,4\}$ both equal the set $\{4\}$.

$\mathbf{F}^n$

To define the higher-dimensional analogues of $\mathbf{R}^2$ and $\mathbf{R}^3$, we will simply replace $\mathbf{R}$ with $\mathbf{F}$ (which equals $\mathbf{R}$ or $\mathbf{C}$) and replace theFana 2 or 3 with an arbitrary positive integer. Specifically, fix a positive integer n for the rest of this section.

1.10 Definition $\mathbf{F}^n$

$\mathbf{F}^n$ is the set of all lists of length n of elements of $\mathbf{F}$:

$$\mathbf{F}^n = \{(x_1,\dots,x_n) : x_j \in \mathbf{F} \text{ for } j = 1,\dots,n\}.$$

For $(x_1,\dots,x_n) \in \mathbf{F}^n$ and $j \in \{1,\dots,n\}$, we say that x_j is the j^{th} ***coordinate*** of $(x_1,\dots,x_n)$.

If $\mathbf{F} = \mathbf{R}$ and n equals 2 or 3, then this definition of $\mathbf{F}^n$ agrees with our previous notions of $\mathbf{R}^2$ and $\mathbf{R}^3$.

1.11 Example $\mathbf{C}^4$ is the set of all lists of four complex numbers:

$$\mathbf{C}^4 = \{(z_1,z_2,z_3,z_4) : z_1,z_2,z_3,z_4 \in \mathbf{C}\}.$$

If $n \ge 4$, we cannot visualize $\mathbf{R}^n$ as a physical object. Similarly, $\mathbf{C}^1$ can be thought of as a plane, but for $n \ge 2$, the human brain cannot provide a full image of $\mathbf{C}^n$. However, even if n is large, we can perform algebraic manipulations in $\mathbf{F}^n$ as easily as in $\mathbf{R}^2$ or $\mathbf{R}^3$. For example, addition in $\mathbf{F}^n$ is defined as follows:

*For an amusing account of how $\mathbf{R}^3$ would be perceived by creatures living in $\mathbf{R}^2$, read **Flatland: A Romance of Many Dimensions**, by Edwin A. Abbott. This novel, published in 1884, may help you imagine a physical space of four or more dimensions.*

1.12 Definition ***addition in*** $\mathbf{F}^n$

Addition in $\mathbf{F}^n$ is defined by adding corresponding coordinates:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n).$$

Often the mathematics of $\mathbf{F}^n$ becomes cleaner if we use a single letter to denote a list of n numbers, without explicitly writing the coordinates. For example, the result below is stated with x and y in $\mathbf{F}^n$ even though the proof requires the more cumbersome notation of $(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)$.

1.13 Commutativity of addition in $\mathbf{F}^n$

If $x, y \in \mathbf{F}^n$, then $x + y = y + x$.

Proof Suppose $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$. Then

$$\begin{aligned} x + y &= (x_1, \dots, x_n) + (y_1, \dots, y_n) \\ &= (x_1 + y_1, \dots, x_n + y_n) \\ &= (y_1 + x_1, \dots, y_n + x_n) \\ &= (y_1, \dots, y_n) + (x_1, \dots, x_n) \\ &= y + x, \end{aligned}$$

where the second and fourth equalities above hold because of the definition of addition in $\mathbf{F}^n$ and the third equality holds because of the usual commutativity of addition in $\mathbf{F}$. ■

The symbol ■ means "end of the proof".

If a single letter is used to denote an element of $\mathbf{F}^n$, then the same letter with appropriate subscripts is often used when coordinates must be displayed. For example, if $x \in \mathbf{F}^n$, then letting x equal $(x_1, \dots, x_n)$ is good notation, as shown in the proof above. Even better, work with just x and avoid explicit coordinates when possible.

1.14 Definition 0

Let 0 denote the list of length n whose coordinates are all 0:

$$0 = (0, \dots, 0).$$

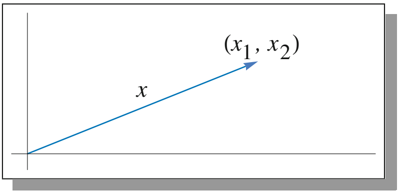
Here we are using the symbol 0 in two different ways—on the left side of the equation in 1.14, the symbol 0 denotes a list of length n, whereas on the right side, each 0 denotes a number. This potentially confusing practice actually causes no problems because the context always makes clear what is intended.

1.15 **Example** Consider the statement that 0 is an additive identity for $\mathbf{F}^n$:

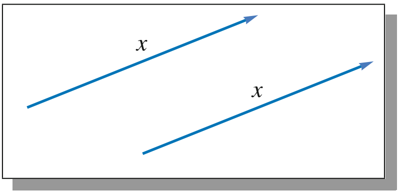
$$x + 0 = x \quad \text{for all } x \in \mathbf{F}^n.$$

Is the 0 above the number 0 or the list 0?

Solution Here 0 is a list, because we have not defined the sum of an element of $\mathbf{F}^n$ (namely, x) and the number 0.



Elements of $\mathbf{R}^2$ can be thought of as points or as vectors.



A vector.

Mathematical models of the economy can have thousands of variables, say $x_1, \dots, x_{5000}$, which means that we must operate in $\mathbf{R}^{5000}$. Such a space cannot be dealt with geometrically. However, the algebraic approach works well. Thus our subject is called ***linear algebra****.*

A picture can aid our intuition. We will draw pictures in $\mathbf{R}^2$ because we can sketch this space on 2-dimensional surfaces such as paper and blackboards. A typical element of $\mathbf{R}^2$ is a point $x = (x_1, x_2)$. Sometimes we think of x not as a point but as an arrow starting at the origin and ending at (x_1, x_2), as shown here. When we think of x as an arrow, we refer to it as a ***vector***.

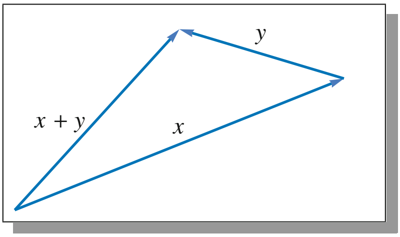
When we think of vectors in $\mathbf{R}^2$ as arrows, we can move an arrow parallel to itself (not changing its length or direction) and still think of it as the same vector. With that viewpoint, you will often gain better understanding by dispensing with the coordinate axes and the explicit coordinates and just thinking of the vector, as shown here.

Whenever we use pictures in $\mathbf{R}^2$ or use the somewhat vague language of points and vectors, remember that these are just aids to our understanding, not substitutes for the actual mathematics that we will develop. Although we cannot draw good pictures in high-dimensional spaces, the elements of these spaces are as rigorously defined as elements of $\mathbf{R}^2$.

For example, $(2, -3, 17, \pi, \sqrt{2})$ is an element of $\mathbf{R}^5$, and we may casually refer to it as a point in $\mathbf{R}^5$ or a vector in $\mathbf{R}^5$ without worrying about whether the geometry of $\mathbf{R}^5$ has any physical meaning.

Recall that we defined the sum of two elements of $\mathbf{F}^n$ to be the element of $\mathbf{F}^n$ obtained by adding corresponding coordinates; see 1.12. As we will now see, addition has a simple geometric interpretation in the special case of $\mathbf{R}^2$.

Suppose we have two vectors x and y in $\mathbf{R}^2$ that we want to add. Move the vector y parallel to itself so that its initial point coincides with the end point of the vector x, as shown here. The sum $x + y$ then equals the vector whose initial point equals the initial point of x and whose end point equals the end point of the vector y, as shown here.



The sum of two vectors.

In the next definition, the 0 on the right side of the displayed equation below is the list $0 \in \mathbf{F}^n$.

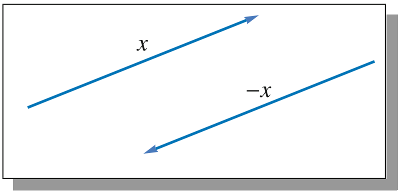
1.16 Definition ***additive inverse in*** $\mathbf{F}^n$

For $x \in \mathbf{F}^n$, the ***additive inverse*** of x, denoted $-x$, is the vector $-x \in \mathbf{F}^n$ such that

$$x + (-x) = 0.$$

In other words, if $x = (x_1, \dots, x_n)$, then $-x = (-x_1, \dots, -x_n)$.

For a vector $x \in \mathbf{R}^2$, the additive inverse $-x$ is the vector parallel to x and with the same length as x but pointing in the opposite direction. The figure here illustrates this way of thinking about the additive inverse in $\mathbf{R}^2$.



A vector and its additive inverse.

Having dealt with addition in $\mathbf{F}^n$, we now turn to multiplication. We could define a multiplication in $\mathbf{F}^n$ in a similar fashion, starting with two elements of $\mathbf{F}^n$ and getting another element of $\mathbf{F}^n$ by multiplying corresponding coordinates. Experience shows that this definition is not useful for our purposes. Another type of multiplication, called scalar multiplication, will be central to our subject. Specifically, we need to define what it means to multiply an element of $\mathbf{F}^n$ by an element of $\mathbf{F}$.

1.17 Definition ***scalar multiplication in $\mathbf{F}^n$***

The ***product*** of a number λ and a vector in $\mathbf{F}^n$ is computed by multiplying each coordinate of the vector by λ:

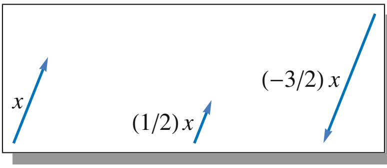
$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n);$$

here $\lambda \in \mathbf{F}$ and $(x_1, \ldots, x_n) \in \mathbf{F}^n$.

In scalar multiplication, we multiply together a scalar and a vector, getting a vector. You may be familiar with the dot product in $\mathbf{R}^2$ or $\mathbf{R}^3$, in which we multiply together two vectors and get a scalar. Generalizations of the dot product will become important when we study inner products in Chapter 6.

Scalar multiplication has a nice geometric interpretation in $\mathbf{R}^2$. If λ is a positive number and x is a vector in $\mathbf{R}^2$, then λx is the vector that points in the same direction as x and whose length is λ times the length of x. In other words, to get λx, we shrink or stretch x by a factor of λ, depending on whether $\lambda < 1$ or $\lambda > 1$.

If λ is a negative number and x is a vector in $\mathbf{R}^2$, then λx is the vector that points in the direction opposite to that of x and whose length is $|\lambda|$ times the length of x, as shown here.



Scalar multiplication.

Digression on Fields

A ***field*** is a set containing at least two distinct elements called 0 and 1, along with operations of addition and multiplication satisfying all the properties listed in 1.3. Thus $\mathbf{R}$ and $\mathbf{C}$ are fields, as is the set of rational numbers along with the usual operations of addition and multiplication. Another example of a field is the set $\{0, 1\}$ with the usual operations of addition and multiplication except that $1 + 1$ is defined to equal 0.

In this book we will not need to deal with fields other than $\mathbf{R}$ and $\mathbf{C}$. However, many of the definitions, theorems, and proofs in linear algebra that work for both $\mathbf{R}$ and $\mathbf{C}$ also work without change for arbitrary fields. If you prefer to do so, throughout Chapters 1, 2, and 3 you can think of $\mathbf{F}$ as denoting an arbitrary field instead of $\mathbf{R}$ or $\mathbf{C}$, except that some of the examples and exercises require that for each positive integer n we have $\underbrace{1 + 1 + \cdots + 1}_{n \text{ times}} \neq 0$.

EXERCISES 1.A

1 Suppose a and b are real numbers, not both 0. Find real numbers c and d such that

$$1/(a+bi) = c + di.$$

2 Show that

$$\frac{-1+\sqrt{3}i}{2}$$

is a cube root of 1 (meaning that its cube equals 1).

3 Find two distinct square roots of i.

4 Show that $\alpha + \beta = \beta + \alpha$ for all $\alpha, \beta \in \mathbf{C}$.

5 Show that $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

6 Show that $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbf{C}$.

7 Show that for every $\alpha \in \mathbf{C}$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha + \beta = 0$.

8 Show that for every $\alpha \in \mathbf{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbf{C}$ such that $\alpha\beta = 1$.

9 Show that $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbf{C}$.

10 Find $x \in \mathbf{R}^4$ such that

$$(4, -3, 1, 7) + 2x = (5, 9, -6, 8).$$

11 Explain why there does not exist $\lambda \in \mathbf{C}$ such that

$$\lambda(2-3i, 5+4i, -6+7i) = (12-5i, 7+22i, -32-9i).$$

12 Show that $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbf{F}^n$.

13 Show that $(ab)x = a(bx)$ for all $x \in \mathbf{F}^n$ and all $a, b \in \mathbf{F}$.

14 Show that $1x = x$ for all $x \in \mathbf{F}^n$.

15 Show that $\lambda(x + y) = \lambda x + \lambda y$ for all $\lambda \in \mathbf{F}$ and all $x, y \in \mathbf{F}^n$.

16 Show that $(a + b)x = ax + bx$ for all $a, b \in \mathbf{F}$ and all $x \in \mathbf{F}^n$.

1.B *Definition of Vector Space*

The motivation for the definition of a vector space comes from properties of addition and scalar multiplication in $\mathbf{F}^n$: Addition is commutative, associative, and has an identity. Every element has an additive inverse. Scalar multiplication is associative. Scalar multiplication by 1 acts as expected. Addition and scalar multiplication are connected by distributive properties.

We will define a vector space to be a set V with an addition and a scalar multiplication on V that satisfy the properties in the paragraph above.

1.18 Definition ***addition, scalar multiplication***

- An ***addition*** on a set V is a function that assigns an element $u+v \in V$ to each pair of elements $u, v \in V$.
- A ***scalar multiplication*** on a set V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbf{F}$ and each $v \in V$.

Now we are ready to give the formal definition of a vector space.

1.19 Definition ***vector space***

A ***vector space*** is a set V along with an addition on V and a scalar multiplication on V such that the following properties hold:

commutativity
$u + v = v + u$ for all $u, v \in V$;

associativity
$(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbf{F}$;

additive identity
there exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$;

additive inverse
for every $v \in V$, there exists $w \in V$ such that $v + w = 0$;

multiplicative identity
$1v = v$ for all $v \in V$;

distributive properties
$a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in \mathbf{F}$ and all $u, v \in V$.

The following geometric language sometimes aids our intuition.

1.20 Definition ***vector, point***

Elements of a vector space are called ***vectors*** or ***points***.

The scalar multiplication in a vector space depends on $\mathbf{F}$. Thus when we need to be precise, we will say that V is a ***vector space over*** $\mathbf{F}$ instead of saying simply that V is a vector space. For example, $\mathbf{R}^n$ is a vector space over $\mathbf{R}$, and $\mathbf{C}^n$ is a vector space over $\mathbf{C}$.

1.21 Definition ***real vector space, complex vector space***

- A vector space over $\mathbf{R}$ is called a ***real vector space***.
- A vector space over $\mathbf{C}$ is called a ***complex vector space***.

Usually the choice of $\mathbf{F}$ is either obvious from the context or irrelevant. Thus we often assume that $\mathbf{F}$ is lurking in the background without specifically mentioning it.

With the usual operations of addition and scalar multiplication, $\mathbf{F}^n$ is a vector space over $\mathbf{F}$, as you should verify. The example of $\mathbf{F}^n$ motivated our definition of vector space.

The simplest vector space contains only one point. In other words, $\{0\}$ is a vector space.

1.22 Example $\mathbf{F}^\infty$ is defined to be the set of all sequences of elements of $\mathbf{F}$:

$$\mathbf{F}^\infty = \{(x_1, x_2, \dots) : x_j \in \mathbf{F} \text{ for } j = 1, 2, \dots\}.$$

Addition and scalar multiplication on $\mathbf{F}^\infty$ are defined as expected:

$$(x_1, x_2, \dots) + (y_1, y_2, \dots) = (x_1 + y_1, x_2 + y_2, \dots),$$
$$\lambda(x_1, x_2, \dots) = (\lambda x_1, \lambda x_2, \dots).$$

With these definitions, $\mathbf{F}^\infty$ becomes a vector space over $\mathbf{F}$, as you should verify. The additive identity in this vector space is the sequence of all 0's.

Our next example of a vector space involves a set of functions.

1.23 Notation $\mathbf{F}^S$

- If S is a set, then $\mathbf{F}^S$ denotes the set of functions from S to $\mathbf{F}$.
- For $f, g \in \mathbf{F}^S$, the ***sum*** $f + g \in \mathbf{F}^S$ is the function defined by
$$(f + g)(x) = f(x) + g(x)$$
for all $x \in S$.
- For $\lambda \in \mathbf{F}$ and $f \in \mathbf{F}^S$, the ***product*** $\lambda f \in \mathbf{F}^S$ is the function defined by
$$(\lambda f)(x) = \lambda f(x)$$
for all $x \in S$.

As an example of the notation above, if S is the interval $[0, 1]$ and $\mathbf{F} = \mathbf{R}$, then $\mathbf{R}^{[0,1]}$ is the set of real-valued functions on the interval $[0, 1]$.

You should verify all three bullet points in the next example.

1.24 Example $\mathbf{F}^S$ ***is a vector space***

- If S is a nonempty set, then $\mathbf{F}^S$ (with the operations of addition and scalar multiplication as defined above) is a vector space over $\mathbf{F}$.
- The additive identity of $\mathbf{F}^S$ is the function $0 : S \to \mathbf{F}$ defined by
$$0(x) = 0$$
for all $x \in S$.
- For $f \in \mathbf{F}^S$, the additive inverse of f is the function $-f : S \to \mathbf{F}$ defined by
$$(-f)(x) = -f(x)$$
for all $x \in S$.

The elements of the vector space $\mathbf{R}^{[0,1]}$ are real-valued functions on $[0, 1]$, not lists. In general, a vector space is an abstract entity whose elements might be lists, functions, or weird objects.

Our previous examples of vector spaces, $\mathbf{F}^n$ and $\mathbf{F}^\infty$, are special cases of the vector space $\mathbf{F}^S$ because a list of length n of numbers in $\mathbf{F}$ can be thought of as a function from $\{1, 2, \dots, n\}$ to $\mathbf{F}$ and a sequence of numbers in $\mathbf{F}$ can be thought of as a function from the set of positive integers to $\mathbf{F}$. In other words, we can think of $\mathbf{F}^n$ as $\mathbf{F}^{\{1,2,\dots,n\}}$ and we can think of $\mathbf{F}^\infty$ as $\mathbf{F}^{\{1,2,\dots\}}$.

Soon we will see further examples of vector spaces, but first we need to develop some of the elementary properties of vector spaces.

The definition of a vector space requires that it have an additive identity. The result below states that this identity is unique.

1.25 Unique additive identity

A vector space has a unique additive identity.

Proof Suppose 0 and $0'$ are both additive identities for some vector space V. Then

$$0' = 0' + 0 = 0 + 0' = 0,$$

where the first equality holds because 0 is an additive identity, the second equality comes from commutativity, and the third equality holds because $0'$ is an additive identity. Thus $0' = 0$, proving that V has only one additive identity. ■

Each element v in a vector space has an additive inverse, an element w in the vector space such that $v + w = 0$. The next result shows that each element in a vector space has only one additive inverse.

1.26 Unique additive inverse

Every element in a vector space has a unique additive inverse.

Proof Suppose V is a vector space. Let $v \in V$. Suppose w and w' are additive inverses of v. Then

$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w'.$$

Thus $w = w'$, as desired. ■

Because additive inverses are unique, the following notation now makes sense.

1.27 Notation $-v$, $w - v$

Let $v, w \in V$. Then

- $-v$ denotes the additive inverse of v;
- $w - v$ is defined to be $w + (-v)$.

Almost all the results in this book involve some vector space. To avoid having to restate frequently that V is a vector space, we now make the necessary declaration once and for all:

1.28 Notation V

For the rest of the book, V denotes a vector space over $\mathbf{F}$.

In the next result, 0 denotes a scalar (the number $0 \in \mathbf{F}$) on the left side of the equation and a vector (the additive identity of V) on the right side of the equation.

1.29 The number 0 times a vector

$0v = 0$ for every $v \in V$.

Note that 1.29 asserts something about scalar multiplication and the additive identity of V. The only part of the definition of a vector space that connects scalar multiplication and vector addition is the distributive property. Thus the distributive property must be used in the proof of 1.29.

Proof For $v \in V$, we have

$$0v = (0 + 0)v = 0v + 0v.$$

Adding the additive inverse of $0v$ to both sides of the equation above gives $0 = 0v$, as desired. ∎

In the next result, 0 denotes the additive identity of V. Although their proofs are similar, 1.29 and 1.30 are not identical. More precisely, 1.29 states that the product of the scalar 0 and any vector equals the vector 0, whereas 1.30 states that the product of any scalar and the vector 0 equals the vector 0.

1.30 A number times the vector 0

$a0 = 0$ for every $a \in \mathbf{F}$.

Proof For $a \in \mathbf{F}$, we have

$$a0 = a(0 + 0) = a0 + a0.$$

Adding the additive inverse of $a0$ to both sides of the equation above gives $0 = a0$, as desired. ∎

Now we show that if an element of V is multiplied by the scalar -1, then the result is the additive inverse of the element of V.

1.31 The number -1 times a vector

$(-1)v = -v$ for every $v \in V$.

Proof For $v \in V$, we have

$$v + (-1)v = 1v + (-1)v = \big(1 + (-1)\big)v = 0v = 0.$$

This equation says that $(-1)v$, when added to v, gives 0. Thus $(-1)v$ is the additive inverse of v, as desired. ■

EXERCISES 1.B

1 Prove that $-(-v) = v$ for every $v \in V$.

2 Suppose $a \in \mathbf{F}$, $v \in V$, and $av = 0$. Prove that $a = 0$ or $v = 0$.

3 Suppose $v, w \in V$. Explain why there exists a unique $x \in V$ such that $v + 3x = w$.

4 The empty set is not a vector space. The empty set fails to satisfy only one of the requirements listed in 1.19. Which one?

5 Show that in the definition of a vector space (1.19), the additive inverse condition can be replaced with the condition that

$$0v = 0 \text{ for all } v \in V.$$

Here the 0 on the left side is the number 0, and the 0 on the right side is the additive identity of V. (The phrase "a condition can be replaced" in a definition means that the collection of objects satisfying the definition is unchanged if the original condition is replaced with the new condition.)

6 Let ∞ and $-\infty$ denote two distinct objects, neither of which is in $\mathbf{R}$. Define an addition and scalar multiplication on $\mathbf{R} \cup \{\infty\} \cup \{-\infty\}$ as you could guess from the notation. Specifically, the sum and product of two real numbers is as usual, and for $t \in \mathbf{R}$ define

$$t\infty = \begin{cases} -\infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ \infty & \text{if } t > 0, \end{cases} \qquad t(-\infty) = \begin{cases} \infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ -\infty & \text{if } t > 0, \end{cases}$$

$$t + \infty = \infty + t = \infty, \qquad t + (-\infty) = (-\infty) + t = -\infty,$$

$$\infty + \infty = \infty, \qquad (-\infty) + (-\infty) = -\infty, \qquad \infty + (-\infty) = 0.$$

Is $\mathbf{R} \cup \{\infty\} \cup \{-\infty\}$ a vector space over $\mathbf{R}$? Explain.

1.C *Subspaces*

By considering subspaces, we can greatly expand our examples of vector spaces.

1.32 Definition ***subspace***

A subset U of V is called a ***subspace*** of V if U is also a vector space (using the same addition and scalar multiplication as on V).

1.33 Example $\{(x_1, x_2, 0) : x_1, x_2 \in \mathbf{F}\}$ is a subspace of $\mathbf{F}^3$.

*Some mathematicians use the term **linear subspace**, which means the same as subspace.*

The next result gives the easiest way to check whether a subset of a vector space is a subspace.

1.34 Conditions for a subspace

A subset U of V is a subspace of V if and only if U satisfies the following three conditions:

additive identity

$0 \in U$

closed under addition

$u, w \in U$ implies $u + w \in U$;

closed under scalar multiplication

$a \in \mathbf{F}$ and $u \in U$ implies $au \in U$.

The additive identity condition above could be replaced with the condition that U is nonempty (then taking $u \in U$, multiplying it by 0, and using the condition that U is closed under scalar multiplication would imply that $0 \in U$). However, if U is indeed a subspace of V, then the easiest way to show that U is nonempty is to show that $0 \in U$.

Proof If U is a subspace of V, then U satisfies the three conditions above by the definition of vector space.

Conversely, suppose U satisfies the three conditions above. The first condition above ensures that the additive identity of V is in U.

The second condition above ensures that addition makes sense on U. The third condition ensures that scalar multiplication makes sense on U.

If $u \in U$, then $-u$ [which equals $(-1)u$ by 1.31] is also in U by the third condition above. Hence every element of U has an additive inverse in U.

The other parts of the definition of a vector space, such as associativity and commutativity, are automatically satisfied for U because they hold on the larger space V. Thus U is a vector space and hence is a subspace of V. ∎

The three conditions in the result above usually enable us to determine quickly whether a given subset of V is a subspace of V. You should verify all the assertions in the next example.

1.35 **Example** ***subspaces***

(a) If $b \in \mathbf{F}$, then

$$\{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_3 = 5x_4 + b\}$$

is a subspace of $\mathbf{F}^4$ if and only if $b = 0$.

(b) The set of continuous real-valued functions on the interval $[0, 1]$ is a subspace of $\mathbf{R}^{[0,1]}$.

(c) The set of differentiable real-valued functions on $\mathbf{R}$ is a subspace of $\mathbf{R}^{\mathbf{R}}$.

(d) The set of differentiable real-valued functions f on the interval $(0, 3)$ such that $f'(2) = b$ is a subspace of $\mathbf{R}^{(0,3)}$ if and only if $b = 0$.

(e) The set of all sequences of complex numbers with limit 0 is a subspace of $\mathbf{C}^\infty$.

Clearly $\{0\}$ is the smallest subspace of V and V itself is the largest subspace of V. The empty set is not a subspace of V because a subspace must be a vector space and hence must contain at least one element, namely, an additive identity.

Verifying some of the items above shows the linear structure underlying parts of calculus. For example, the second item above requires the result that the sum of two continuous functions is continuous. As another example, the fourth item above requires the result that for a constant c, the derivative of cf equals c times the derivative of f.

The subspaces of $\mathbf{R}^2$ are precisely $\{0\}$, $\mathbf{R}^2$, and all lines in $\mathbf{R}^2$ through the origin. The subspaces of $\mathbf{R}^3$ are precisely $\{0\}$, $\mathbf{R}^3$, all lines in $\mathbf{R}^3$ through the origin, and all planes in $\mathbf{R}^3$ through the origin. To prove that all these objects are indeed subspaces is easy—the hard part is to show that they are the only subspaces of $\mathbf{R}^2$ and $\mathbf{R}^3$. That task will be easier after we introduce some additional tools in the next chapter.

Sums of Subspaces

The union of subspaces is rarely a subspace (see Exercise 12), which is why we usually work with sums rather than unions.

When dealing with vector spaces, we are usually interested only in subspaces, as opposed to arbitrary subsets. The notion of the sum of subspaces will be useful.

1.36 Definition ***sum of subsets***

Suppose $U_1, \dots, U_m$ are subsets of V. The ***sum*** of $U_1, \dots, U_m$, denoted $U_1 + \cdots + U_m$, is the set of all possible sums of elements of $U_1, \dots, U_m$. More precisely,

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \dots, u_m \in U_m\}.$$

Let's look at some examples of sums of subspaces.

1.37 Example Suppose U is the set of all elements of $\mathbf{F}^3$ whose second and third coordinates equal 0, and W is the set of all elements of $\mathbf{F}^3$ whose first and third coordinates equal 0:

$$U = \{(x, 0, 0) \in \mathbf{F}^3 : x \in \mathbf{F}\} \quad \text{and} \quad W = \{(0, y, 0) \in \mathbf{F}^3 : y \in \mathbf{F}\}.$$

Then

$$U + W = \{(x, y, 0) : x, y \in \mathbf{F}\},$$

as you should verify.

1.38 Example Suppose that $U = \{(x, x, y, y) \in \mathbf{F}^4 : x, y \in \mathbf{F}\}$ and $W = \{(x, x, x, y) \in \mathbf{F}^4 : x, y \in \mathbf{F}\}$. Then

$$U + W = \{(x, x, y, z) \in \mathbf{F}^4 : x, y, z \in \mathbf{F}\},$$

as you should verify.

The next result states that the sum of subspaces is a subspace, and is in fact the smallest subspace containing all the summands.

1.39 Sum of subspaces is the smallest containing subspace

Suppose $U_1, \dots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, \dots, U_m$.

Proof It is easy to see that $0 \in U_1 + \cdots + U_m$ and that $U_1 + \cdots + U_m$ is closed under addition and scalar multiplication. Thus 1.34 implies that $U_1 + \cdots + U_m$ is a subspace of V.

Clearly $U_1, \ldots, U_m$ are all contained in $U_1 + \cdots + U_m$ (to see this, consider sums $u_1 + \cdots + u_m$ where all except one of the u's are 0). Conversely, every subspace of V containing $U_1, \ldots, U_m$ contains $U_1 + \cdots + U_m$ (because subspaces must contain all finite sums of their elements). Thus $U_1 + \cdots + U_m$ is the smallest subspace of V containing $U_1, \ldots, U_m$. ■

Sums of subspaces in the theory of vector spaces are analogous to unions of subsets in set theory. Given two subspaces of a vector space, the smallest subspace containing them is their sum. Analogously, given two subsets of a set, the smallest subset containing them is their union.

Direct Sums

Suppose $U_1, \ldots, U_m$ are subspaces of V. Every element of $U_1 + \cdots + U_m$ can be written in the form

$$u_1 + \cdots + u_m,$$

where each u_j is in U_j. We will be especially interested in cases where each vector in $U_1 + \cdots + U_m$ can be represented in the form above in only one way. This situation is so important that we give it a special name: direct sum.

1.40 Definition ***direct sum***

Suppose $U_1, \ldots, U_m$ are subspaces of V.

- The sum $U_1 + \cdots + U_m$ is called a ***direct sum*** if each element of $U_1 + \cdots + U_m$ can be written in only one way as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j.
- If $U_1 + \cdots + U_m$ is a direct sum, then $U_1 \oplus \cdots \oplus U_m$ denotes $U_1 + \cdots + U_m$, with the $\oplus$ notation serving as an indication that this is a direct sum.

1.41 Example Suppose U is the subspace of $\mathbf{F}^3$ of those vectors whose last coordinate equals 0, and W is the subspace of $\mathbf{F}^3$ of those vectors whose first two coordinates equal 0:

$$U = \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\} \quad \text{and} \quad W = \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\}.$$

Then $\mathbf{F}^3 = U \oplus W$, as you should verify.

1.42 Example Suppose U_j is the subspace of $\mathbf{F}^n$ of those vectors whose coordinates are all 0, except possibly in the j^{th} slot (thus, for example, $U_2 = \{(0, x, 0, \ldots, 0) \in \mathbf{F}^n : x \in \mathbf{F}\}$). Then

$$\mathbf{F}^n = U_1 \oplus \cdots \oplus U_n,$$

as you should verify.

Sometimes nonexamples add to our understanding as much as examples.

1.43 Example Let

$$\begin{aligned} U_1 &= \{(x, y, 0) \in \mathbf{F}^3 : x, y \in \mathbf{F}\}, \\ U_2 &= \{(0, 0, z) \in \mathbf{F}^3 : z \in \mathbf{F}\}, \\ U_3 &= \{(0, y, y) \in \mathbf{F}^3 : y \in \mathbf{F}\}. \end{aligned}$$

Show that $U_1 + U_2 + U_3$ is not a direct sum.

Solution Clearly $\mathbf{F}^3 = U_1 + U_2 + U_3$, because every vector $(x, y, z) \in \mathbf{F}^3$ can be written as

$$(x, y, z) = (x, y, 0) + (0, 0, z) + (0, 0, 0),$$

where the first vector on the right side is in U_1, the second vector is in U_2, and the third vector is in U_3.

However, $\mathbf{F}^3$ does not equal the direct sum of U_1, U_2, U_3, because the vector $(0, 0, 0)$ can be written in two different ways as a sum $u_1 + u_2 + u_3$, with each u_j in U_j. Specifically, we have

$$(0, 0, 0) = (0, 1, 0) + (0, 0, 1) + (0, -1, -1)$$

and, of course,

$$(0, 0, 0) = (0, 0, 0) + (0, 0, 0) + (0, 0, 0),$$

where the first vector on the right side of each equation above is in U_1, the second vector is in U_2, and the third vector is in U_3.

The symbol $\oplus$, which is a plus sign inside a circle, serves as a reminder that we are dealing with a special type of sum of subspaces—each element in the direct sum can be represented only one way as a sum of elements from the specified subspaces.

The definition of direct sum requires that every vector in the sum have a unique representation as an appropriate sum. The next result shows that when deciding whether a sum of subspaces is a direct sum, we need only consider whether 0 can be uniquely written as an appropriate sum.

1.44 Condition for a direct sum

Suppose $U_1, \dots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0.

Proof First suppose $U_1 + \cdots + U_m$ is a direct sum. Then the definition of direct sum implies that the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0.

Now suppose that the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0. To show that $U_1 + \cdots + U_m$ is a direct sum, let $v \in U_1 + \cdots + U_m$. We can write

$$v = u_1 + \cdots + u_m$$

for some $u_1 \in U_1, \dots, u_m \in U_m$. To show that this representation is unique, suppose we also have

$$v = v_1 + \cdots + v_m,$$

where $v_1 \in U_1, \dots, v_m \in U_m$. Subtracting these two equations, we have

$$0 = (u_1 - v_1) + \cdots + (u_m - v_m).$$

Because $u_1 - v_1 \in U_1, \dots, u_m - v_m \in U_m$, the equation above implies that each $u_j - v_j$ equals 0. Thus $u_1 = v_1, \dots, u_m = v_m$, as desired. ■

The next result gives a simple condition for testing which pairs of subspaces give a direct sum.

1.45 Direct sum of two subspaces

Suppose U and W are subspaces of V. Then $U + W$ is a direct sum if and only if $U \cap W = \{0\}$.

Proof First suppose that $U + W$ is a direct sum. If $v \in U \cap W$, then $0 = v + (-v)$, where $v \in U$ and $-v \in W$. By the unique representation of 0 as the sum of a vector in U and a vector in W, we have $v = 0$. Thus $U \cap W = \{0\}$, completing the proof in one direction.

To prove the other direction, now suppose $U \cap W = \{0\}$. To prove that $U + W$ is a direct sum, suppose $u \in U$, $w \in W$, and

$$0 = u + w.$$

To complete the proof, we need only show that $u = w = 0$ (by 1.44). The equation above implies that $u = -w \in W$. Thus $u \in U \cap W$. Hence $u = 0$, which by the equation above implies that $w = 0$, completing the proof. ■

Sums of subspaces are analogous to unions of subsets. Similarly, direct sums of subspaces are analogous to disjoint unions of subsets. No two subspaces of a vector space can be disjoint, because both contain 0*. So disjointness is replaced, at least in the case of two subspaces, with the requirement that the intersection equals* $\{0\}$.

The result above deals only with the case of two subspaces. When asking about a possible direct sum with more than two subspaces, it is not enough to test that each pair of the subspaces intersect only at 0. To see this, consider Example 1.43. In that nonexample of a direct sum, we have $U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{0\}$.

EXERCISES 1.C

1 For each of the following subsets of $\mathbf{F}^3$, determine whether it is a subspace of $\mathbf{F}^3$:

(a) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 0\}$;

(b) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 + 2x_2 + 3x_3 = 4\}$;

(c) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 x_2 x_3 = 0\}$;

(d) $\{(x_1, x_2, x_3) \in \mathbf{F}^3 : x_1 = 5x_3\}$.

2 Verify all the assertions in Example 1.35.

3 Show that the set of differentiable real-valued functions f on the interval $(-4, 4)$ such that $f'(-1) = 3f(2)$ is a subspace of $\mathbf{R}^{(-4,4)}$.

4 Suppose $b \in \mathbf{R}$. Show that the set of continuous real-valued functions f on the interval $[0, 1]$ such that $\int_0^1 f = b$ is a subspace of $\mathbf{R}^{[0,1]}$ if and only if $b = 0$.

5 Is $\mathbf{R}^2$ a subspace of the complex vector space $\mathbf{C}^2$?

6 (a) Is $\{(a, b, c) \in \mathbf{R}^3 : a^3 = b^3\}$ a subspace of $\mathbf{R}^3$?

(b) Is $\{(a, b, c) \in \mathbf{C}^3 : a^3 = b^3\}$ a subspace of $\mathbf{C}^3$?

7 Give an example of a nonempty subset U of $\mathbf{R}^2$ such that U is closed under addition and under taking additive inverses (meaning $-u \in U$ whenever $u \in U$), but U is not a subspace of $\mathbf{R}^2$.

8 Give an example of a nonempty subset U of $\mathbf{R}^2$ such that U is closed under scalar multiplication, but U is not a subspace of $\mathbf{R}^2$.

9 A function $f: \mathbf{R} \to \mathbf{R}$ is called ***periodic*** if there exists a positive number p such that $f(x) = f(x + p)$ for all $x \in \mathbf{R}$. Is the set of periodic functions from $\mathbf{R}$ to $\mathbf{R}$ a subspace of $\mathbf{R}^{\mathbf{R}}$? Explain.

10 Suppose U_1 and U_2 are subspaces of V. Prove that the intersection $U_1 \cap U_2$ is a subspace of V.

11 Prove that the intersection of every collection of subspaces of V is a subspace of V.

12 Prove that the union of two subspaces of V is a subspace of V if and only if one of the subspaces is contained in the other.

13 Prove that the union of three subspaces of V is a subspace of V if and only if one of the subspaces contains the other two.
[*This exercise is surprisingly harder than the previous exercise, possibly because this exercise is not true if we replace* $\mathbf{F}$ *with a field containing only two elements.*]

14 Verify the assertion in Example 1.38.

15 Suppose U is a subspace of V. What is $U + U$?

16 Is the operation of addition on the subspaces of V commutative? In other words, if U and W are subspaces of V, is $U + W = W + U$?

17 Is the operation of addition on the subspaces of V associative? In other words, if U_1, U_2, U_3 are subspaces of V, is

$$(U_1 + U_2) + U_3 = U_1 + (U_2 + U_3)?$$

18 Does the operation of addition on the subspaces of V have an additive identity? Which subspaces have additive inverses?

19 Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that

$$U_1 + W = U_2 + W,$$

then $U_1 = U_2$.

20 Suppose

$$U = \{(x, x, y, y) \in \mathbf{F}^4 : x, y \in \mathbf{F}\}.$$

Find a subspace W of $\mathbf{F}^4$ such that $\mathbf{F}^4 = U \oplus W$.

21 Suppose

$$U = \{(x, y, x + y, x - y, 2x) \in \mathbf{F}^5 : x, y \in \mathbf{F}\}.$$

Find a subspace W of $\mathbf{F}^5$ such that $\mathbf{F}^5 = U \oplus W$.

22 Suppose

$$U = \{(x, y, x + y, x - y, 2x) \in \mathbf{F}^5 : x, y \in \mathbf{F}\}.$$

Find three subspaces W_1, W_2, W_3 of $\mathbf{F}^5$, none of which equals $\{0\}$, such that $\mathbf{F}^5 = U \oplus W_1 \oplus W_2 \oplus W_3$.

23 Prove or give a counterexample: if U_1, U_2, W are subspaces of V such that

$$V = U_1 \oplus W \quad \text{and} \quad V = U_2 \oplus W,$$

then $U_1 = U_2$.

24 A function $f : \mathbf{R} \to \mathbf{R}$ is called ***even*** if

$$f(-x) = f(x)$$

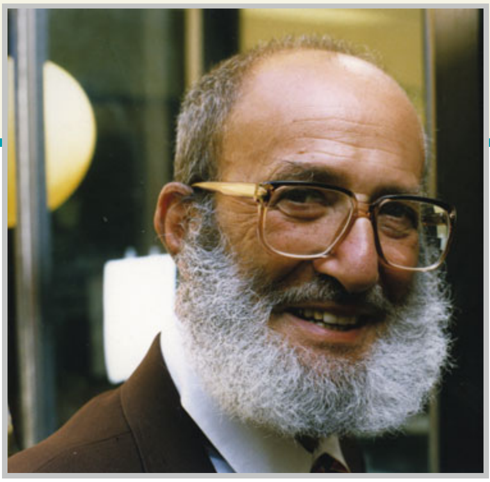
for all $x \in \mathbf{R}$. A function $f : \mathbf{R} \to \mathbf{R}$ is called ***odd*** if

$$f(-x) = -f(x)$$

for all $x \in \mathbf{R}$. Let U_{e} denote the set of real-valued even functions on $\mathbf{R}$ and let U_{o} denote the set of real-valued odd functions on $\mathbf{R}$. Show that $\mathbf{R}^{\mathbf{R}} = U_{\mathrm{e}} \oplus U_{\mathrm{o}}$.

CHAPTER

2



American mathematician Paul Halmos (1916–2006), who in 1942 published the first modern linear algebra book. The title of Halmos's book was the same as the title of this chapter.

Finite-Dimensional Vector Spaces

Let's review our standing assumptions:

2.1 Notation F, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a vector space over $\mathbf{F}$.

In the last chapter we learned about vector spaces. Linear algebra focuses not on arbitrary vector spaces, but on finite-dimensional vector spaces, which we introduce in this chapter.

LEARNING OBJECTIVES FOR THIS CHAPTER

- span
- linear independence
- bases
- dimension

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_2

2.A Span and Linear Independence

We have been writing lists of numbers surrounded by parentheses, and we will continue to do so for elements of $\mathbf{F}^n$; for example, $(2, -7, 8) \in \mathbf{F}^3$. However, now we need to consider lists of vectors (which may be elements of $\mathbf{F}^n$ or of other vector spaces). To avoid confusion, we will usually write lists of vectors without surrounding parentheses. For example, $(4, 1, 6), (9, 5, 7)$ is a list of length 2 of vectors in $\mathbf{R}^3$.

2.2 Notation ***list of vectors***

We will usually write lists of vectors without surrounding parentheses.

Linear Combinations and Span

Adding up scalar multiples of vectors in a list gives what is called a linear combination of the list. Here is the formal definition:

2.3 Definition ***linear combination***

A ***linear combination*** of a list $v_1, \dots, v_m$ of vectors in V is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, \dots, a_m \in \mathbf{F}$.

2.4 Example In $\mathbf{F}^3$,

- $(17, -4, 2)$ is a linear combination of $(2, 1, -3), (1, -2, 4)$ because

$$(17, -4, 2) = 6(2, 1, -3) + 5(1, -2, 4).$$

- $(17, -4, 5)$ is not a linear combination of $(2, 1, -3), (1, -2, 4)$ because there do not exist numbers $a_1, a_2 \in \mathbf{F}$ such that

$$(17, -4, 5) = a_1(2, 1, -3) + a_2(1, -2, 4).$$

  In other words, the system of equations

$$\begin{aligned} 17 &= 2a_1 + a_2 \\ -4 &= a_1 - 2a_2 \\ 5 &= -3a_1 + 4a_2 \end{aligned}$$

  has no solutions (as you should verify).

2.5 Definition ***span***

The set of all linear combinations of a list of vectors $v_1, \ldots, v_m$ in V is called the ***span*** of $v_1, \ldots, v_m$, denoted $\operatorname{span}(v_1, \ldots, v_m)$. In other words,

$$\operatorname{span}(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : a_1, \ldots, a_m \in \mathbf{F}\}.$$

The span of the empty list $(\ )$ is defined to be $\{0\}$.

2.6 Example The previous example shows that in $\mathbf{F}^3$,

- $(17, -4, 2) \in \operatorname{span}\big((2, 1, -3), (1, -2, 4)\big)$;
- $(17, -4, 5) \notin \operatorname{span}\big((2, 1, -3), (1, -2, 4)\big)$.

Some mathematicians use the term ***linear span***, which means the same as span.

2.7 Span is the smallest containing subspace

The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.

Proof Suppose $v_1, \ldots, v_m$ is a list of vectors in V.

First we show that $\operatorname{span}(v_1, \ldots, v_m)$ is a subspace of V. The additive identity is in $\operatorname{span}(v_1, \ldots, v_m)$, because

$$0 = 0v_1 + \cdots + 0v_m.$$

Also, $\operatorname{span}(v_1, \ldots, v_m)$ is closed under addition, because

$$(a_1 v_1 + \cdots + a_m v_m) + (c_1 v_1 + \cdots + c_m v_m) = (a_1 + c_1) v_1 + \cdots + (a_m + c_m) v_m.$$

Furthermore, $\operatorname{span}(v_1, \ldots, v_m)$ is closed under scalar multiplication, because

$$\lambda(a_1 v_1 + \cdots + a_m v_m) = \lambda a_1 v_1 + \cdots + \lambda a_m v_m.$$

Thus $\operatorname{span}(v_1, \ldots, v_m)$ is a subspace of V (by 1.34).

Each v_j is a linear combination of $v_1, \ldots, v_m$ (to show this, set $a_j = 1$ and let the other a's in 2.3 equal 0). Thus $\operatorname{span}(v_1, \ldots, v_m)$ contains each v_j. Conversely, because subspaces are closed under scalar multiplication and addition, every subspace of V containing each v_j contains $\operatorname{span}(v_1, \ldots, v_m)$. Thus $\operatorname{span}(v_1, \ldots, v_m)$ is the smallest subspace of V containing all the vectors $v_1, \ldots, v_m$. ■

2.8 Definition ***spans***

If $\operatorname{span}(v_1, \ldots, v_m)$ equals V, we say that $v_1, \ldots, v_m$ ***spans*** V.

2.9 Example Suppose n is a positive integer. Show that

$$(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$$

spans $\mathbf{F}^n$. Here the j^{th} vector in the list above is the n-tuple with 1 in the j^{th} slot and 0 in all other slots.

Solution Suppose $(x_1, \ldots, x_n) \in \mathbf{F}^n$. Then

$$(x_1, \ldots, x_n) = x_1(1, 0, \ldots, 0) + x_2(0, 1, 0, \ldots, 0) + \cdots + x_n(0, \ldots, 0, 1).$$

Thus $(x_1, \ldots, x_n) \in \operatorname{span}\big((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)\big)$, as desired.

Now we can make one of the key definitions in linear algebra.

2.10 Definition ***finite-dimensional vector space***

A vector space is called ***finite-dimensional*** if some list of vectors in it spans the space.

Recall that by definition every list has finite length.

Example 2.9 above shows that $\mathbf{F}^n$ is a finite-dimensional vector space for every positive integer n.

The definition of a polynomial is no doubt already familiar to you.

2.11 Definition ***polynomial,*** $\mathcal{P}(\mathbf{F})$

- A function $p\colon \mathbf{F} \to \mathbf{F}$ is called a ***polynomial*** with coefficients in $\mathbf{F}$ if there exist $a_0, \ldots, a_m \in \mathbf{F}$ such that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbf{F}$.

- $\mathcal{P}(\mathbf{F})$ is the set of all polynomials with coefficients in $\mathbf{F}$.

With the usual operations of addition and scalar multiplication, $\mathcal{P}(\mathbf{F})$ is a vector space over $\mathbf{F}$, as you should verify. In other words, $\mathcal{P}(\mathbf{F})$ is a subspace of $\mathbf{F}^{\mathbf{F}}$, the vector space of functions from $\mathbf{F}$ to $\mathbf{F}$.

If a polynomial (thought of as a function from $\mathbf{F}$ to $\mathbf{F}$) is represented by two sets of coefficients, then subtracting one representation of the polynomial from the other produces a polynomial that is identically zero as a function on $\mathbf{F}$ and hence has all zero coefficients (if you are unfamiliar with this fact, just believe it for now; we will prove it later—see 4.7). **Conclusion:** the coefficients of a polynomial are uniquely determined by the polynomial. Thus the next definition uniquely defines the degree of a polynomial.

2.12 Definition ***degree of a polynomial,*** $\deg p$

- A polynomial $p \in \mathcal{P}(\mathbf{F})$ is said to have ***degree*** m if there exist scalars $a_0, a_1, \ldots, a_m \in \mathbf{F}$ with $a_m \neq 0$ such that

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

  for all $z \in \mathbf{F}$. If p has degree m, we write $\deg p = m$.

- The polynomial that is identically 0 is said to have degree $-\infty$.

In the next definition, we use the convention that $-\infty < m$, which means that the polynomial 0 is in $\mathcal{P}_m(\mathbf{F})$.

2.13 Definition $\mathcal{P}_m(\mathbf{F})$

For m a nonnegative integer, $\mathcal{P}_m(\mathbf{F})$ denotes the set of all polynomials with coefficients in $\mathbf{F}$ and degree at most m.

To verify the next example, note that $\mathcal{P}_m(\mathbf{F}) = \operatorname{span}(1, z, \ldots, z^m)$; here we are slightly abusing notation by letting z^k denote a function.

2.14 Example $\mathcal{P}_m(\mathbf{F})$ is a finite-dimensional vector space for each nonnegative integer m.

2.15 Definition ***infinite-dimensional vector space***

A vector space is called ***infinite-dimensional*** if it is not finite-dimensional.

2.16 Example Show that $\mathcal{P}(\mathbf{F})$ is infinite-dimensional.

Solution Consider any list of elements of $\mathcal{P}(\mathbf{F})$. Let m denote the highest degree of the polynomials in this list. Then every polynomial in the span of this list has degree at most m. Thus z^{m+1} is not in the span of our list. Hence no list spans $\mathcal{P}(\mathbf{F})$. Thus $\mathcal{P}(\mathbf{F})$ is infinite-dimensional.

Linear Independence

Suppose $v_1, \dots, v_m \in V$ and $v \in \operatorname{span}(v_1, \dots, v_m)$. By the definition of span, there exist $a_1, \dots, a_m \in \mathbf{F}$ such that

$$v = a_1 v_1 + \cdots + a_m v_m.$$

Consider the question of whether the choice of scalars in the equation above is unique. Suppose $c_1, \dots, c_m$ is another set of scalars such that

$$v = c_1 v_1 + \cdots + c_m v_m.$$

Subtracting the last two equations, we have

$$0 = (a_1 - c_1) v_1 + \cdots + (a_m - c_m) v_m.$$

Thus we have written 0 as a linear combination of $(v_1, \dots, v_m)$. If the only way to do this is the obvious way (using 0 for all scalars), then each $a_j - c_j$ equals 0, which means that each a_j equals c_j (and thus the choice of scalars was indeed unique). This situation is so important that we give it a special name—linear independence—which we now define.

2.17 Definition ***linearly independent***

- A list $v_1, \dots, v_m$ of vectors in V is called ***linearly independent*** if the only choice of $a_1, \dots, a_m \in \mathbf{F}$ that makes $a_1 v_1 + \cdots + a_m v_m$ equal 0 is $a_1 = \cdots = a_m = 0$.
- The empty list $(\,)$ is also declared to be linearly independent.

The reasoning above shows that $v_1, \dots, v_m$ is linearly independent if and only if each vector in $\operatorname{span}(v_1, \dots, v_m)$ has only one representation as a linear combination of $v_1, \dots, v_m$.

2.18 Example ***linearly independent lists***

(a) A list v of one vector $v \in V$ is linearly independent if and only if $v \neq 0$.

(b) A list of two vectors in V is linearly independent if and only if neither vector is a scalar multiple of the other.

(c) $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$ is linearly independent in $\mathbf{F}^4$.

(d) The list $1, z, \dots, z^m$ is linearly independent in $\mathcal{P}(\mathbf{F})$ for each nonnegative integer m.

If some vectors are removed from a linearly independent list, the remaining list is also linearly independent, as you should verify.

2.19 Definition ***linearly dependent***

- A list of vectors in V is called ***linearly dependent*** if it is not linearly independent.
- In other words, a list $v_1, \dots, v_m$ of vectors in V is linearly dependent if there exist $a_1, \dots, a_m \in \mathbf{F}$, not all 0, such that $a_1 v_1 + \dots + a_m v_m = 0$.

2.20 Example ***linearly dependent lists***

- $(2, 3, 1), (1, -1, 2), (7, 3, 8)$ is linearly dependent in $\mathbf{F}^3$ because
$$2(2, 3, 1) + 3(1, -1, 2) + (-1)(7, 3, 8) = (0, 0, 0).$$
- The list $(2, 3, 1), (1, -1, 2), (7, 3, c)$ is linearly dependent in $\mathbf{F}^3$ if and only if $c = 8$, as you should verify.
- If some vector in a list of vectors in V is a linear combination of the other vectors, then the list is linearly dependent. (Proof: After writing one vector in the list as equal to a linear combination of the other vectors, move that vector to the other side of the equation, where it will be multiplied by -1.)
- Every list of vectors in V containing the 0 vector is linearly dependent. (This is a special case of the previous bullet point.)

The lemma below will often be useful. It states that given a linearly dependent list of vectors, one of the vectors is in the span of the previous ones and furthermore we can throw out that vector without changing the span of the original list.

2.21 Linear Dependence Lemma

Suppose $v_1, \ldots, v_m$ is a linearly dependent list in V. Then there exists $j \in \{1, 2, \ldots, m\}$ such that the following hold:

(a) $v_j \in \operatorname{span}(v_1, \ldots, v_{j-1})$;

(b) if the j^{th} term is removed from $v_1, \ldots, v_m$, the span of the remaining list equals $\operatorname{span}(v_1, \ldots, v_m)$.

Proof Because the list $v_1, \ldots, v_m$ is linearly dependent, there exist numbers $a_1, \ldots, a_m \in \mathbf{F}$, not all 0, such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Let j be the largest element of $\{1, \ldots, m\}$ such that $a_j \neq 0$. Then

2.22
$$v_j = -\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1},$$

proving (a).

To prove (b), suppose $u \in \operatorname{span}(v_1, \ldots, v_m)$. Then there exist numbers $c_1, \ldots, c_m \in \mathbf{F}$ such that

$$u = c_1 v_1 + \cdots + c_m v_m.$$

In the equation above, we can replace v_j with the right side of 2.22, which shows that u is in the span of the list obtained by removing the j^{th} term from $v_1, \ldots, v_m$. Thus (b) holds. ■

Choosing $j = 1$ in the Linear Dependence Lemma above means that $v_1 = 0$, because if $j = 1$ then condition (a) above is interpreted to mean that $v_1 \in \operatorname{span}(\,)$; recall that $\operatorname{span}(\,) = \{0\}$. Note also that the proof of part (b) above needs to be modified in an obvious way if $v_1 = 0$ and $j = 1$.

In general, the proofs in the rest of the book will not call attention to special cases that must be considered involving empty lists, lists of length 1, the subspace $\{0\}$, or other trivial cases for which the result is clearly true but needs a slightly different proof. Be sure to check these special cases yourself.

Now we come to a key result. It says that no linearly independent list in V is longer than a spanning list in V.

2.23 Length of linearly independent list $\leq$ length of spanning list

In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

Proof Suppose $u_1, \dots, u_m$ is linearly independent in V. Suppose also that $w_1, \dots, w_n$ spans V. We need to prove that $m \leq n$. We do so through the multi-step process described below; note that in each step we add one of the u's and remove one of the w's.

Step 1

Let B be the list $w_1, \dots, w_n$, which spans V. Thus adjoining any vector in V to this list produces a linearly dependent list (because the newly adjoined vector can be written as a linear combination of the other vectors). In particular, the list

$$u_1, w_1, \dots, w_n$$

is linearly dependent. Thus by the Linear Dependence Lemma (2.21), we can remove one of the w's so that the new list B (of length n) consisting of u_1 and the remaining w's spans V.

Step j

The list B (of length n) from step $j - 1$ spans V. Thus adjoining any vector to this list produces a linearly dependent list. In particular, the list of length $(n + 1)$ obtained by adjoining u_j to B, placing it just after $u_1, \dots, u_{j-1}$, is linearly dependent. By the Linear Dependence Lemma (2.21), one of the vectors in this list is in the span of the previous ones, and because $u_1, \dots, u_j$ is linearly independent, this vector is one of the w's, not one of the u's. We can remove that w from B so that the new list B (of length n) consisting of $u_1, \dots, u_j$ and the remaining w's spans V.

After step m, we have added all the u's and the process stops. At each step as we add a u to B, the Linear Dependence Lemma implies that there is some w to remove. Thus there are at least as many w's as u's. ■

The next two examples show how the result above can be used to show, without any computations, that certain lists are not linearly independent and that certain lists do not span a given vector space.

2.24 Example Show that the list $(1,2,3), (4,5,8), (9,6,7), (-3,2,8)$ is not linearly independent in $\mathbf{R}^3$.

Solution The list $(1,0,0), (0,1,0), (0,0,1)$ spans $\mathbf{R}^3$. Thus no list of length larger than 3 is linearly independent in $\mathbf{R}^3$.

2.25 Example Show that the list $(1,2,3,-5), (4,5,8,3), (9,6,7,-1)$ does not span $\mathbf{R}^4$.

Solution The list $(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$ is linearly independent in $\mathbf{R}^4$. Thus no list of length less than 4 spans $\mathbf{R}^4$.

Our intuition suggests that every subspace of a finite-dimensional vector space should also be finite-dimensional. We now prove that this intuition is correct.

2.26 Finite-dimensional subspaces

Every subspace of a finite-dimensional vector space is finite-dimensional.

Proof Suppose V is finite-dimensional and U is a subspace of V. We need to prove that U is finite-dimensional. We do this through the following multi-step construction.

Step 1

If $U = \{0\}$, then U is finite-dimensional and we are done. If $U \neq \{0\}$, then choose a nonzero vector $v_1 \in U$.

Step j

If $U = \operatorname{span}(v_1, \dots, v_{j-1})$, then U is finite-dimensional and we are done. If $U \neq \operatorname{span}(v_1, \dots, v_{j-1})$, then choose a vector $v_j \in U$ such that

$$v_j \notin \operatorname{span}(v_1, \dots, v_{j-1}).$$

After each step, as long as the process continues, we have constructed a list of vectors such that no vector in this list is in the span of the previous vectors. Thus after each step we have constructed a linearly independent list, by the Linear Dependence Lemma (2.21). This linearly independent list cannot be longer than any spanning list of V (by 2.23). Thus the process eventually terminates, which means that U is finite-dimensional. ■

EXERCISES 2.A

1 Suppose v_1, v_2, v_3, v_4 spans V. Prove that the list

$$v_1 - v_2, v_2 - v_3, v_3 - v_4, v_4$$

also spans V.

2 Verify the assertions in Example 2.18.

3 Find a number t such that

$$(3, 1, 4), (2, -3, 5), (5, 9, t)$$

is not linearly independent in $\mathbf{R}^3$.

4 Verify the assertion in the second bullet point in Example 2.20.

5 (a) Show that if we think of $\mathbf{C}$ as a vector space over $\mathbf{R}$, then the list $(1 + i, 1 - i)$ is linearly independent.

(b) Show that if we think of $\mathbf{C}$ as a vector space over $\mathbf{C}$, then the list $(1 + i, 1 - i)$ is linearly dependent.

6 Suppose v_1, v_2, v_3, v_4 is linearly independent in V. Prove that the list

$$v_1 - v_2, v_2 - v_3, v_3 - v_4, v_4$$

is also linearly independent.

7 Prove or give a counterexample: If $v_1, v_2, \ldots, v_m$ is a linearly independent list of vectors in V, then

$$5v_1 - 4v_2, v_2, v_3, \ldots, v_m$$

is linearly independent.

8 Prove or give a counterexample: If $v_1, v_2, \ldots, v_m$ is a linearly independent list of vectors in V and $\lambda \in \mathbf{F}$ with $\lambda \neq 0$, then $\lambda v_1, \lambda v_2, \ldots, \lambda v_m$ is linearly independent.

9 Prove or give a counterexample: If $v_1, \ldots, v_m$ and $w_1, \ldots, w_m$ are linearly independent lists of vectors in V, then $v_1 + w_1, \ldots, v_m + w_m$ is linearly independent.

10 Suppose $v_1, \ldots, v_m$ is linearly independent in V and $w \in V$. Prove that if $v_1 + w, \ldots, v_m + w$ is linearly dependent, then $w \in \operatorname{span}(v_1, \ldots, v_m)$.

11 Suppose $v_1, \dots, v_m$ is linearly independent in V and $w \in V$. Show that $v_1, \dots, v_m, w$ is linearly independent if and only if

$$w \notin \operatorname{span}(v_1, \dots, v_m).$$

12 Explain why there does not exist a list of six polynomials that is linearly independent in $\mathcal{P}_4(\mathbf{F})$.

13 Explain why no list of four polynomials spans $\mathcal{P}_4(\mathbf{F})$.

14 Prove that V is infinite-dimensional if and only if there is a sequence $v_1, v_2, \dots$ of vectors in V such that $v_1, \dots, v_m$ is linearly independent for every positive integer m.

15 Prove that $\mathbf{F}^\infty$ is infinite-dimensional.

16 Prove that the real vector space of all continuous real-valued functions on the interval $[0, 1]$ is infinite-dimensional.

17 Suppose $p_0, p_1, \dots, p_m$ are polynomials in $\mathcal{P}_m(\mathbf{F})$ such that $p_j(2) = 0$ for each j. Prove that $p_0, p_1, \dots, p_m$ is not linearly independent in $\mathcal{P}_m(\mathbf{F})$.

2.B *Bases*

In the last section, we discussed linearly independent lists and spanning lists. Now we bring these concepts together.

2.27 Definition ***basis***

A ***basis*** of V is a list of vectors in V that is linearly independent and spans V.

2.28 Example ***bases***

(a) The list $(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$ is a basis of $\mathbf{F}^n$, called the ***standard basis*** of $\mathbf{F}^n$.

(b) The list $(1, 2), (3, 5)$ is a basis of $\mathbf{F}^2$.

(c) The list $(1, 2, -4), (7, -5, 6)$ is linearly independent in $\mathbf{F}^3$ but is not a basis of $\mathbf{F}^3$ because it does not span $\mathbf{F}^3$.

(d) The list $(1, 2), (3, 5), (4, 13)$ spans $\mathbf{F}^2$ but is not a basis of $\mathbf{F}^2$ because it is not linearly independent.

(e) The list $(1, 1, 0), (0, 0, 1)$ is a basis of $\{(x, x, y) \in \mathbf{F}^3 : x, y \in \mathbf{F}\}$.

(f) The list $(1, -1, 0), (1, 0, -1)$ is a basis of

$$\{(x, y, z) \in \mathbf{F}^3 : x + y + z = 0\}.$$

(g) The list $1, z, \ldots, z^m$ is a basis of $\mathcal{P}_m(\mathbf{F})$.

In addition to the standard basis, $\mathbf{F}^n$ has many other bases. For example, $(7, 5), (-4, 9)$ and $(1, 2), (3, 5)$ are both bases of $\mathbf{F}^2$.

The next result helps explain why bases are useful. Recall that "uniquely" means "in only one way".

2.29 Criterion for basis

A list $v_1, \ldots, v_n$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form

2.30
$$v = a_1 v_1 + \cdots + a_n v_n,$$

where $a_1, \ldots, a_n \in \mathbf{F}$.

Proof First suppose that $v_1, \dots, v_n$ is a basis of V. Let $v \in V$. Because $v_1, \dots, v_n$ spans V, there exist $a_1, \dots, a_n \in \mathbf{F}$ such that 2.30 holds. To show that the representation in 2.30 is unique, suppose $c_1, \dots, c_n$ are scalars such that we also have

This proof is essentially a repetition of the ideas that led us to the definition of linear independence.

$$v = c_1 v_1 + \cdots + c_n v_n.$$

Subtracting the last equation from 2.30, we get

$$0 = (a_1 - c_1)v_1 + \cdots + (a_n - c_n)v_n.$$

This implies that each $a_j - c_j$ equals 0 (because $v_1, \dots, v_n$ is linearly independent). Hence $a_1 = c_1, \dots, a_n = c_n$. We have the desired uniqueness, completing the proof in one direction.

For the other direction, suppose every $v \in V$ can be written uniquely in the form given by 2.30. Clearly this implies that $v_1, \dots, v_n$ spans V. To show that $v_1, \dots, v_n$ is linearly independent, suppose $a_1, \dots, a_n \in \mathbf{F}$ are such that

$$0 = a_1 v_1 + \cdots + a_n v_n.$$

The uniqueness of the representation 2.30 (taking $v = 0$) now implies that $a_1 = \cdots = a_n = 0$. Thus $v_1, \dots, v_n$ is linearly independent and hence is a basis of V. ■

A spanning list in a vector space may not be a basis because it is not linearly independent. Our next result says that given any spanning list, some (possibly none) of the vectors in it can be discarded so that the remaining list is linearly independent and still spans the vector space.

As an example in the vector space $\mathbf{F}^2$, if the procedure in the proof below is applied to the list $(1,2), (3,6), (4,7), (5,9)$, then the second and fourth vectors will be removed. This leaves $(1,2), (4,7)$, which is a basis of $\mathbf{F}^2$.

2.31 Spanning list contains a basis

Every spanning list in a vector space can be reduced to a basis of the vector space.

Proof Suppose $v_1, \dots, v_n$ spans V. We want to remove some of the vectors from $v_1, \dots, v_n$ so that the remaining vectors form a basis of V. We do this through the multi-step process described below.

Start with B equal to the list $v_1, \dots, v_n$.

Step 1

If $v_1 = 0$, delete v_1 from B. If $v_1 \neq 0$, leave B unchanged.

Step j

If v_j is in $\operatorname{span}(v_1, \dots, v_{j-1})$, delete v_j from B. If v_j is not in $\operatorname{span}(v_1, \dots, v_{j-1})$, leave B unchanged.

Stop the process after step n, getting a list B. This list B spans V because our original list spanned V and we have discarded only vectors that were already in the span of the previous vectors. The process ensures that no vector in B is in the span of the previous ones. Thus B is linearly independent, by the Linear Dependence Lemma (2.21). Hence B is a basis of V. ■

Our next result, an easy corollary of the previous result, tells us that every finite-dimensional vector space has a basis.

2.32 Basis of finite-dimensional vector space

Every finite-dimensional vector space has a basis.

Proof By definition, a finite-dimensional vector space has a spanning list. The previous result tells us that each spanning list can be reduced to a basis. ■

Our next result is in some sense a dual of 2.31, which said that every spanning list can be reduced to a basis. Now we show that given any linearly independent list, we can adjoin some additional vectors (this includes the possibility of adjoining no additional vectors) so that the extended list is still linearly independent but also spans the space.

2.33 Linearly independent list extends to a basis

Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.

Proof Suppose $u_1, \dots, u_m$ is linearly independent in a finite-dimensional vector space V. Let $w_1, \dots, w_n$ be a basis of V. Thus the list

$$u_1, \dots, u_m, w_1, \dots, w_n$$

spans V. Applying the procedure of the proof of 2.31 to reduce this list to a basis of V produces a basis consisting of the vectors $u_1, \dots, u_m$ (none of the u's get deleted in this procedure because $u_1, \dots, u_m$ is linearly independent) and some of the w's. ■

As an example in $\mathbf{F}^3$, suppose we start with the linearly independent list $(2,3,4),(9,6,8)$. If we take w_1, w_2, w_3 in the proof above to be the standard basis of $\mathbf{F}^3$, then the procedure in the proof above produces the list $(2,3,4),(9,6,8),(0,1,0)$, which is a basis of $\mathbf{F}^3$.

Using the same basic ideas but considerably more advanced tools, the next result can be proved without the hypothesis that V is finite-dimensional.

As an application of the result above, we now show that every subspace of a finite-dimensional vector space can be paired with another subspace to form a direct sum of the whole space.

2.34 Every subspace of V is part of a direct sum equal to V

Suppose V is finite-dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.

Proof Because V is finite-dimensional, so is U (see 2.26). Thus there is a basis $u_1, \dots, u_m$ of U (see 2.32). Of course $u_1, \dots, u_m$ is a linearly independent list of vectors in V. Hence this list can be extended to a basis $u_1, \dots, u_m, w_1, \dots, w_n$ of V (see 2.33). Let $W = \operatorname{span}(w_1, \dots, w_n)$.

To prove that $V = U \oplus W$, by 1.45 we need only show that

$$V = U + W \quad \text{and} \quad U \cap W = \{0\}.$$

To prove the first equation above, suppose $v \in V$. Then, because the list $u_1, \dots, u_m, w_1, \dots, w_n$ spans V, there exist $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbf{F}$ such that

$$v = \underbrace{a_1 u_1 + \cdots + a_m u_m}_{u} + \underbrace{b_1 w_1 + \cdots + b_n w_n}_{w}.$$

In other words, we have $v = u + w$, where $u \in U$ and $w \in W$ are defined as above. Thus $v \in U + W$, completing the proof that $V = U + W$.

To show that $U \cap W = \{0\}$, suppose $v \in U \cap W$. Then there exist scalars $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbf{F}$ such that

$$v = a_1 u_1 + \cdots + a_m u_m = b_1 w_1 + \cdots + b_n w_n.$$

Thus

$$a_1 u_1 + \cdots + a_m u_m - b_1 w_1 - \cdots - b_n w_n = 0.$$

Because $u_1, \dots, u_m, w_1, \dots, w_n$ is linearly independent, this implies that $a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0$. Thus $v = 0$, completing the proof that $U \cap W = \{0\}$. ■

EXERCISES 2.B

1 Find all vector spaces that have exactly one basis.

2 Verify all the assertions in Example 2.28.

3 (a) Let U be the subspace of $\mathbf{R}^5$ defined by

$$U = \{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{R}^5 : x_1 = 3x_2 \text{ and } x_3 = 7x_4\}.$$

Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathbf{R}^5$.

(c) Find a subspace W of $\mathbf{R}^5$ such that $\mathbf{R}^5 = U \oplus W$.

4 (a) Let U be the subspace of $\mathbf{C}^5$ defined by

$$U = \{(z_1, z_2, z_3, z_4, z_5) \in \mathbf{C}^5 : 6z_1 = z_2 \text{ and } z_3 + 2z_4 + 3z_5 = 0\}.$$

Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathbf{C}^5$.

(c) Find a subspace W of $\mathbf{C}^5$ such that $\mathbf{C}^5 = U \oplus W$.

5 Prove or disprove: there exists a basis p_0, p_1, p_2, p_3 of $\mathcal{P}_3(\mathbf{F})$ such that none of the polynomials p_0, p_1, p_2, p_3 has degree 2.

6 Suppose v_1, v_2, v_3, v_4 is a basis of V. Prove that

$$v_1 + v_2, v_2 + v_3, v_3 + v_4, v_4$$

is also a basis of V.

7 Prove or give a counterexample: If v_1, v_2, v_3, v_4 is a basis of V and U is a subspace of V such that $v_1, v_2 \in U$ and $v_3 \notin U$ and $v_4 \notin U$, then v_1, v_2 is a basis of U.

8 Suppose U and W are subspaces of V such that $V = U \oplus W$. Suppose also that $u_1, \dots, u_m$ is a basis of U and $w_1, \dots, w_n$ is a basis of W. Prove that

$$u_1, \dots, u_m, w_1, \dots, w_n$$

is a basis of V.

2.C Dimension

Although we have been discussing finite-dimensional vector spaces, we have not yet defined the dimension of such an object. How should dimension be defined? A reasonable definition should force the dimension of $\mathbf{F}^n$ to equal n. Notice that the standard basis

$$(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$$

of $\mathbf{F}^n$ has length n. Thus we are tempted to define the dimension as the length of a basis. However, a finite-dimensional vector space in general has many different bases, and our attempted definition makes sense only if all bases in a given vector space have the same length. Fortunately that turns out to be the case, as we now show.

2.35 Basis length does not depend on basis

Any two bases of a finite-dimensional vector space have the same length.

Proof Suppose V is finite-dimensional. Let B_1 and B_2 be two bases of V. Then B_1 is linearly independent in V and B_2 spans V, so the length of B_1 is at most the length of B_2 (by 2.23). Interchanging the roles of B_1 and B_2, we also see that the length of B_2 is at most the length of B_1. Thus the length of B_1 equals the length of B_2, as desired. ■

Now that we know that any two bases of a finite-dimensional vector space have the same length, we can formally define the dimension of such spaces.

2.36 Definition ***dimension*, dim V**

- The ***dimension*** of a finite-dimensional vector space is the length of any basis of the vector space.
- The dimension of V (if V is finite-dimensional) is denoted by $\dim V$.

2.37 Example ***dimensions***

- $\dim \mathbf{F}^n = n$ because the standard basis of $\mathbf{F}^n$ has length n.
- $\dim \mathcal{P}_m(\mathbf{F}) = m + 1$ because the basis $1, z, \ldots, z^m$ of $\mathcal{P}_m(\mathbf{F})$ has length $m + 1$.

Every subspace of a finite-dimensional vector space is finite-dimensional (by 2.26) and so has a dimension. The next result gives the expected inequality about the dimension of a subspace.

2.38 Dimension of a subspace

If V is finite-dimensional and U is a subspace of V, then $\dim U \leq \dim V$.

Proof Suppose V is finite-dimensional and U is a subspace of V. Think of a basis of U as a linearly independent list in V, and think of a basis of V as a spanning list in V. Now use 2.23 to conclude that $\dim U \leq \dim V$. ■

To check that a list of vectors in V is a basis of V, we must, according to the definition, show that the list in question satisfies two properties: it must be linearly independent and it must span V. The next two results show that if the list in question has the right length, then we need only check that it satisfies one of the two required properties. First we prove that every linearly independent list with the right length is a basis.

The real vector space $\mathbf{R}^2$ has dimension 2; the complex vector space $\mathbf{C}$ has dimension 1. As sets, $\mathbf{R}^2$ can be identified with $\mathbf{C}$ (and addition is the same on both spaces, as is scalar multiplication by real numbers). Thus when we talk about the dimension of a vector space, the role played by the choice of $\mathbf{F}$ cannot be neglected.

2.39 Linearly independent list of the right length is a basis

Suppose V is finite-dimensional. Then every linearly independent list of vectors in V with length $\dim V$ is a basis of V.

Proof Suppose $\dim V = n$ and $v_1, \dots, v_n$ is linearly independent in V. The list $v_1, \dots, v_n$ can be extended to a basis of V (by 2.33). However, every basis of V has length n, so in this case the extension is the trivial one, meaning that no elements are adjoined to $v_1, \dots, v_n$. In other words, $v_1, \dots, v_n$ is a basis of V, as desired. ■

2.40 Example Show that the list $(5, 7), (4, 3)$ is a basis of $\mathbf{F}^2$.

Solution This list of two vectors in $\mathbf{F}^2$ is obviously linearly independent (because neither vector is a scalar multiple of the other). Note that $\mathbf{F}^2$ has dimension 2. Thus 2.39 implies that the linearly independent list $(5, 7), (4, 3)$ of length 2 is a basis of $\mathbf{F}^2$ (we do not need to bother checking that it spans $\mathbf{F}^2$).

2.41 Example Show that $1, (x-5)^2, (x-5)^3$ is a basis of the subspace U of $\mathcal{P}_3(\mathbf{R})$ defined by

$$U = \{p \in \mathcal{P}_3(\mathbf{R}) : p'(5) = 0\}.$$

Solution Clearly each of the polynomials 1, $(x-5)^2$, and $(x-5)^3$ is in U.

Suppose $a, b, c \in \mathbf{R}$ and

$$a + b(x-5)^2 + c(x-5)^3 = 0$$

for every $x \in \mathbf{R}$. Without explicitly expanding the left side of the equation above, we can see that the left side has a cx^3 term. Because the right side has no x^3 term, this implies that $c = 0$. Because $c = 0$, we see that the left side has a bx^2 term, which implies that $b = 0$. Because $b = c = 0$, we can also conclude that $a = 0$.

Thus the equation above implies that $a = b = c = 0$. Hence the list $1, (x-5)^2, (x-5)^3$ is linearly independent in U.

Thus $\dim U \geq 3$. Because U is a subspace of $\mathcal{P}_3(\mathbf{R})$, we know that $\dim U \leq \dim \mathcal{P}_3(\mathbf{R}) = 4$ (by 2.38). However, $\dim U$ cannot equal 4, because otherwise when we extend a basis of U to a basis of $\mathcal{P}_3(\mathbf{R})$ we would get a list with length greater than 4. Hence $\dim U = 3$. Thus 2.39 implies that the linearly independent list $1, (x-5)^2, (x-5)^3$ is a basis of U.

Now we prove that a spanning list with the right length is a basis.

2.42 Spanning list of the right length is a basis

Suppose V is finite-dimensional. Then every spanning list of vectors in V with length $\dim V$ is a basis of V.

Proof Suppose $\dim V = n$ and $v_1, \dots, v_n$ spans V. The list $v_1, \dots, v_n$ can be reduced to a basis of V (by 2.31). However, every basis of V has length n, so in this case the reduction is the trivial one, meaning that no elements are deleted from $v_1, \dots, v_n$. In other words, $v_1, \dots, v_n$ is a basis of V, as desired. ■

The next result gives a formula for the dimension of the sum of two subspaces of a finite-dimensional vector space. This formula is analogous to a familiar counting formula: the number of elements in the union of two finite sets equals the number of elements in the first set, plus the number of elements in the second set, minus the number of elements in the intersection of the two sets.

2.43 Dimension of a sum

If U_1 and U_2 are subspaces of a finite-dimensional vector space, then

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Proof Let $u_1, \dots, u_m$ be a basis of $U_1 \cap U_2$; thus $\dim(U_1 \cap U_2) = m$. Because $u_1, \dots, u_m$ is a basis of $U_1 \cap U_2$, it is linearly independent in U_1. Hence this list can be extended to a basis $u_1, \dots, u_m, v_1, \dots, v_j$ of U_1 (by 2.33). Thus $\dim U_1 = m + j$. Also extend $u_1, \dots, u_m$ to a basis $u_1, \dots, u_m, w_1, \dots, w_k$ of U_2; thus $\dim U_2 = m + k$.

We will show that

$$u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k$$

is a basis of $U_1 + U_2$. This will complete the proof, because then we will have

$$\begin{aligned}\dim(U_1 + U_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).\end{aligned}$$

Clearly $\operatorname{span}(u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k)$ contains U_1 and U_2 and hence equals $U_1 + U_2$. So to show that this list is a basis of $U_1 + U_2$ we need only show that it is linearly independent. To prove this, suppose

$$a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j + c_1w_1 + \cdots + c_kw_k = 0,$$

where all the a's, b's, and c's are scalars. We need to prove that all the a's, b's, and c's equal 0. The equation above can be rewritten as

$$c_1w_1 + \cdots + c_kw_k = -a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j,$$

which shows that $c_1w_1 + \cdots + c_kw_k \in U_1$. All the w's are in U_2, so this implies that $c_1w_1 + \cdots + c_kw_k \in U_1 \cap U_2$. Because $u_1, \dots, u_m$ is a basis of $U_1 \cap U_2$, we can write

$$c_1w_1 + \cdots + c_kw_k = d_1u_1 + \cdots + d_mu_m$$

for some choice of scalars $d_1, \dots, d_m$. But $u_1, \dots, u_m, w_1, \dots, w_k$ is linearly independent, so the last equation implies that all the c's (and d's) equal 0. Thus our original equation involving the a's, b's, and c's becomes

$$a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j = 0.$$

Because the list $u_1, \dots, u_m, v_1, \dots, v_j$ is linearly independent, this equation implies that all the a's and b's are 0. We now know that all the a's, b's, and c's equal 0, as desired. ■

EXERCISES 2.C

1 Suppose V is finite-dimensional and U is a subspace of V such that $\dim U = \dim V$. Prove that $U = V$.

2 Show that the subspaces of $\mathbf{R}^2$ are precisely $\{0\}$, $\mathbf{R}^2$, and all lines in $\mathbf{R}^2$ through the origin.

3 Show that the subspaces of $\mathbf{R}^3$ are precisely $\{0\}$, $\mathbf{R}^3$, all lines in $\mathbf{R}^3$ through the origin, and all planes in $\mathbf{R}^3$ through the origin.

4 (a) Let $U = \{p \in \mathcal{P}_4(\mathbf{F}) : p(6) = 0\}$. Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathcal{P}_4(\mathbf{F})$.

(c) Find a subspace W of $\mathcal{P}_4(\mathbf{F})$ such that $\mathcal{P}_4(\mathbf{F}) = U \oplus W$.

5 (a) Let $U = \{p \in \mathcal{P}_4(\mathbf{R}) : p''(6) = 0\}$. Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathcal{P}_4(\mathbf{R})$.

(c) Find a subspace W of $\mathcal{P}_4(\mathbf{R})$ such that $\mathcal{P}_4(\mathbf{R}) = U \oplus W$.

6 (a) Let $U = \{p \in \mathcal{P}_4(\mathbf{F}) : p(2) = p(5)\}$. Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathcal{P}_4(\mathbf{F})$.

(c) Find a subspace W of $\mathcal{P}_4(\mathbf{F})$ such that $\mathcal{P}_4(\mathbf{F}) = U \oplus W$.

7 (a) Let $U = \{p \in \mathcal{P}_4(\mathbf{F}) : p(2) = p(5) = p(6)\}$. Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathcal{P}_4(\mathbf{F})$.

(c) Find a subspace W of $\mathcal{P}_4(\mathbf{F})$ such that $\mathcal{P}_4(\mathbf{F}) = U \oplus W$.

8 (a) Let $U = \{p \in \mathcal{P}_4(\mathbf{R}) : \int_{-1}^{1} p = 0\}$. Find a basis of U.

(b) Extend the basis in part (a) to a basis of $\mathcal{P}_4(\mathbf{R})$.

(c) Find a subspace W of $\mathcal{P}_4(\mathbf{R})$ such that $\mathcal{P}_4(\mathbf{R}) = U \oplus W$.

9 Suppose $v_1, \dots, v_m$ is linearly independent in V and $w \in V$. Prove that

$$\dim \operatorname{span}(v_1 + w, \dots, v_m + w) \geq m - 1.$$

10 Suppose $p_0, p_1, \dots, p_m \in \mathcal{P}(\mathbf{F})$ are such that each p_j has degree j. Prove that $p_0, p_1, \dots, p_m$ is a basis of $\mathcal{P}_m(\mathbf{F})$.

11 Suppose that U and W are subspaces of $\mathbf{R}^8$ such that $\dim U = 3$, $\dim W = 5$, and $U + W = \mathbf{R}^8$. Prove that $\mathbf{R}^8 = U \oplus W$.

12 Suppose U and W are both five-dimensional subspaces of $\mathbf{R}^9$. Prove that $U \cap W \neq \{0\}$.

13 Suppose U and W are both 4-dimensional subspaces of $\mathbf{C}^6$. Prove that there exist two vectors in $U \cap W$ such that neither of these vectors is a scalar multiple of the other.

14 Suppose $U_1, \dots, U_m$ are finite-dimensional subspaces of V. Prove that $U_1 + \dots + U_m$ is finite-dimensional and

$$\dim(U_1 + \dots + U_m) \leq \dim U_1 + \dots + \dim U_m.$$

15 Suppose V is finite-dimensional, with $\dim V = n \geq 1$. Prove that there exist 1-dimensional subspaces $U_1, \dots, U_n$ of V such that

$$V = U_1 \oplus \dots \oplus U_n.$$

16 Suppose $U_1, \dots, U_m$ are finite-dimensional subspaces of V such that $U_1 + \dots + U_m$ is a direct sum. Prove that $U_1 \oplus \dots \oplus U_m$ is finite-dimensional and

$$\dim U_1 \oplus \dots \oplus U_m = \dim U_1 + \dots + \dim U_m.$$

[*The exercise above deepens the analogy between direct sums of subspaces and disjoint unions of subsets. Specifically, compare this exercise to the following obvious statement: if a set is written as a disjoint union of finite subsets, then the number of elements in the set equals the sum of the numbers of elements in the disjoint subsets.*]

17 You might guess, by analogy with the formula for the number of elements in the union of three subsets of a finite set, that if U_1, U_2, U_3 are subspaces of a finite-dimensional vector space, then

$$\begin{aligned}\dim(U_1 &+ U_2 + U_3)\\ &= \dim U_1 + \dim U_2 + \dim U_3\\ &\quad - \dim(U_1 \cap U_2) - \dim(U_1 \cap U_3) - \dim(U_2 \cap U_3)\\ &\quad + \dim(U_1 \cap U_2 \cap U_3).\end{aligned}$$

Prove this or give a counterexample.

CHAPTER

3

German mathematician Carl Friedrich Gauss (1777–1855), who in 1809 published a method for solving systems of linear equations. This method, now called Gaussian elimination, was also used in a Chinese book published over 1600 years earlier.

Linear Maps

So far our attention has focused on vector spaces. No one gets excited about vector spaces. The interesting part of linear algebra is the subject to which we now turn—linear maps.

In this chapter we will frequently need another vector space, which we will call W, in addition to V. Thus our standing assumptions are now as follows:

3.1 Notation $\mathbf{F}, V, W$

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V and W denote vector spaces over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- Fundamental Theorem of Linear Maps
- the matrix of a linear map with respect to given bases
- isomorphic vector spaces
- product spaces
- quotient spaces
- the dual space of a vector space and the dual of a linear map

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_3

3.A The Vector Space of Linear Maps

Definition and Examples of Linear Maps

Now we are ready for one of the key definitions in linear algebra.

3.2 Definition ***linear map***

A ***linear map*** from V to W is a function $T: V \to W$ with the following properties:

additivity

$T(u+v) = Tu + Tv$ for all $u, v \in V$;

homogeneity

$T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbf{F}$ and all $v \in V$.

*Some mathematicians use the term **linear transformation**, which means the same as linear map.*

Note that for linear maps we often use the notation Tv as well as the more standard functional notation $T(v)$.

3.3 Notation $\mathcal{L}(V, W)$

The set of all linear maps from V to W is denoted $\mathcal{L}(V, W)$.

Let's look at some examples of linear maps. Make sure you verify that each of the functions defined below is indeed a linear map:

3.4 Example ***linear maps***

zero

In addition to its other uses, we let the symbol 0 denote the function that takes each element of some vector space to the additive identity of another vector space. To be specific, $0 \in \mathcal{L}(V, W)$ is defined by

$$0v = 0.$$

The 0 on the left side of the equation above is a function from V to W, whereas the 0 on the right side is the additive identity in W. As usual, the context should allow you to distinguish between the many uses of the symbol 0.

identity

The ***identity map***, denoted I, is the function on some vector space that takes each element to itself. To be specific, $I \in \mathcal{L}(V, V)$ is defined by

$$Iv = v.$$

differentiation
Define $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ by

$$Dp = p'.$$

The assertion that this function is a linear map is another way of stating a basic result about differentiation: $(f + g)' = f' + g'$ and $(\lambda f)' = \lambda f'$ whenever f, g are differentiable and λ is a constant.

integration
Define $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathbf{R}\big)$ by

$$Tp = \int_0^1 p(x)\, dx.$$

The assertion that this function is linear is another way of stating a basic result about integration: the integral of the sum of two functions equals the sum of the integrals, and the integral of a constant times a function equals the constant times the integral of the function.

multiplication by x^2
Define $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ by

$$(Tp)(x) = x^2 p(x)$$

for $x \in \mathbf{R}$.

backward shift
Recall that $\mathbf{F}^\infty$ denotes the vector space of all sequences of elements of $\mathbf{F}$. Define $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$ by

$$T(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots).$$

from $\mathbf{R}^3$ to $\mathbf{R}^2$
Define $T \in \mathcal{L}(\mathbf{R}^3, \mathbf{R}^2)$ by

$$T(x, y, z) = (2x - y + 3z, 7x + 5y - 6z).$$

from $\mathbf{F}^n$ to $\mathbf{F}^m$
Generalizing the previous example, let m and n be positive integers, let $A_{j,k} \in \mathbf{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$, and define $T \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$ by

$$T(x_1, \dots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \dots, A_{m,1}x_1 + \cdots + A_{m,n}x_n).$$

Actually every linear map from $\mathbf{F}^n$ to $\mathbf{F}^m$ is of this form.

The existence part of the next result means that we can find a linear map that takes on whatever values we wish on the vectors in a basis. The uniqueness part of the next result means that a linear map is completely determined by its values on a basis.

3.5 Linear maps and basis of domain

Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_n \in W$. Then there exists a unique linear map $T: V \to W$ such that

$$Tv_j = w_j$$

for each $j = 1, \ldots, n$.

Proof First we show the existence of a linear map T with the desired property. Define $T: V \to W$ by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n,$$

where $c_1, \ldots, c_n$ are arbitrary elements of $\mathbf{F}$. The list $v_1, \ldots, v_n$ is a basis of V, and thus the equation above does indeed define a function T from V to W (because each element of V can be uniquely written in the form $c_1 v_1 + \cdots + c_n v_n$).

For each j, taking $c_j = 1$ and the other c's equal to 0 in the equation above shows that $Tv_j = w_j$.

If $u, v \in V$ with $u = a_1 v_1 + \cdots + a_n v_n$ and $v = c_1 v_1 + \cdots + c_n v_n$, then

$$\begin{aligned} T(u + v) &= T\big((a_1 + c_1)v_1 + \cdots + (a_n + c_n)v_n\big) \\ &= (a_1 + c_1)w_1 + \cdots + (a_n + c_n)w_n \\ &= (a_1 w_1 + \cdots + a_n w_n) + (c_1 w_1 + \cdots + c_n w_n) \\ &= Tu + Tv. \end{aligned}$$

Similarly, if $\lambda \in \mathbf{F}$ and $v = c_1 v_1 + \cdots + c_n v_n$, then

$$\begin{aligned} T(\lambda v) &= T(\lambda c_1 v_1 + \cdots + \lambda c_n v_n) \\ &= \lambda c_1 w_1 + \cdots + \lambda c_n w_n \\ &= \lambda(c_1 w_1 + \cdots + c_n w_n) \\ &= \lambda T v. \end{aligned}$$

Thus T is a linear map from V to W.

To prove uniqueness, now suppose that $T \in \mathcal{L}(V, W)$ and that $Tv_j = w_j$ for $j = 1, \ldots, n$. Let $c_1, \ldots, c_n \in \mathbf{F}$. The homogeneity of T implies that $T(c_j v_j) = c_j w_j$ for $j = 1, \ldots, n$. The additivity of T now implies that

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n.$$

Thus T is uniquely determined on $\operatorname{span}(v_1, \ldots, v_n)$ by the equation above. Because $v_1, \ldots, v_n$ is a basis of V, this implies that T is uniquely determined on V. ■

Algebraic Operations on $\mathcal{L}(V, W)$

We begin by defining addition and scalar multiplication on $\mathcal{L}(V, W)$.

3.6 Definition ***addition and scalar multiplication on*** $\mathcal{L}(V, W)$

Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbf{F}$. The ***sum*** $S + T$ and the ***product*** λT are the linear maps from V to W defined by

$$(S + T)(v) = Sv + Tv \quad \text{and} \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$.

You should verify that $S+T$ and λT as defined above are indeed linear maps. In other words, if $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbf{F}$, then $S + T \in \mathcal{L}(V, W)$ and $\lambda T \in \mathcal{L}(V, W)$.

Although linear maps are pervasive throughout mathematics, they are not as ubiquitous as imagined by some confused students who seem to think that $\cos$ *is a linear map from* $\mathbf{R}$ *to* $\mathbf{R}$ *when they write that* $\cos 2x$ *equals* $2\cos x$ *and that* $\cos(x + y)$ *equals* $\cos x + \cos y$.

Because we took the trouble to define addition and scalar multiplication on $\mathcal{L}(V, W)$, the next result should not be a surprise.

3.7 $\mathcal{L}(V, W)$ **is a vector space**

With the operations of addition and scalar multiplication as defined above, $\mathcal{L}(V, W)$ is a vector space.

The routine proof of the result above is left to the reader. Note that the additive identity of $\mathcal{L}(V, W)$ is the zero linear map defined earlier in this section.

Usually it makes no sense to multiply together two elements of a vector space, but for some pairs of linear maps a useful product exists. We will need a third vector space, so for the rest of this section suppose U is a vector space over $\mathbf{F}$.

3.8 Definition ***Product of Linear Maps***

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the ***product*** $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for $u \in U$.

In other words, ST is just the usual composition $S \circ T$ of two functions, but when both functions are linear, most mathematicians write ST instead of $S \circ T$. You should verify that ST is indeed a linear map from U to W whenever $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$.

Note that ST is defined only when T maps into the domain of S.

3.9 Algebraic properties of products of linear maps

associativity

$$(T_1 T_2) T_3 = T_1 (T_2 T_3)$$

whenever T_1, T_2, and T_3 are linear maps such that the products make sense (meaning that T_3 maps into the domain of T_2, and T_2 maps into the domain of T_1).

identity

$$TI = IT = T$$

whenever $T \in \mathcal{L}(V, W)$ (the first I is the identity map on V, and the second I is the identity map on W).

distributive properties

$$(S_1 + S_2)T = S_1 T + S_2 T \quad \text{and} \quad S(T_1 + T_2) = ST_1 + ST_2$$

whenever $T, T_1, T_2 \in \mathcal{L}(U, V)$ and $S, S_1, S_2 \in \mathcal{L}(V, W)$.

The routine proof of the result above is left to the reader.

Multiplication of linear maps is not commutative. In other words, it is not necessarily true that $ST = TS$, even if both sides of the equation make sense.

3.10 **Example** Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the differentiation map defined in Example 3.4 and $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the multiplication by x^2 map defined earlier in this section. Show that $TD \neq DT$.

Solution We have

$$\big((TD)p\big)(x) = x^2 p'(x) \quad \text{but} \quad \big((DT)p\big)(x) = x^2 p'(x) + 2xp(x).$$

In other words, differentiating and then multiplying by x^2 is not the same as multiplying by x^2 and then differentiating.

3.11 Linear maps take 0 to 0

Suppose T is a linear map from V to W. Then $T(0) = 0$.

Proof By additivity, we have

$$T(0) = T(0 + 0) = T(0) + T(0).$$

Add the additive inverse of $T(0)$ to each side of the equation above to conclude that $T(0) = 0$. ■

EXERCISES 3.A

1 Suppose $b, c \in \mathbf{R}$. Define $T : \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x, y, z) = (2x - 4y + 3z + b, 6x + cxyz).$$

Show that T is linear if and only if $b = c = 0$.

2 Suppose $b, c \in \mathbf{R}$. Define $T : \mathcal{P}(\mathbf{R}) \to \mathbf{R}^2$ by

$$Tp = \left(3p(4) + 5p'(6) + bp(1)p(2), \int_{-1}^{2} x^3 p(x)\, dx + c \sin p(0)\right).$$

Show that T is linear if and only if $b = c = 0$.

3 Suppose $T \in \mathcal{L}(\mathbf{F}^n, \mathbf{F}^m)$. Show that there exist scalars $A_{j,k} \in \mathbf{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$ such that

$$T(x_1, \dots, x_n) = (A_{1,1}x_1 + \dots + A_{1,n}x_n, \dots, A_{m,1}x_1 + \dots + A_{m,n}x_n)$$

for every $(x_1, \dots, x_n) \in \mathbf{F}^n$.
[*The exercise above shows that T has the form promised in the last item of Example 3.4.*]

4 Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_m$ is a list of vectors in V such that $Tv_1, \dots, Tv_m$ is a linearly independent list in W. Prove that $v_1, \dots, v_m$ is linearly independent.

5 Prove the assertion in 3.7.

6 Prove the assertions in 3.9.

7 Show that every linear map from a 1-dimensional vector space to itself is multiplication by some scalar. More precisely, prove that if $\dim V = 1$ and $T \in \mathcal{L}(V, V)$, then there exists $\lambda \in \mathbf{F}$ such that $Tv = \lambda v$ for all $v \in V$.

8 Give an example of a function $\varphi\colon \mathbf{R}^2 \to \mathbf{R}$ such that

$$\varphi(av) = a\varphi(v)$$

for all $a \in \mathbf{R}$ and all $v \in \mathbf{R}^2$ but φ is not linear.
[*The exercise above and the next exercise show that neither homogeneity nor additivity alone is enough to imply that a function is a linear map.*]

9 Give an example of a function $\varphi\colon \mathbf{C} \to \mathbf{C}$ such that

$$\varphi(w + z) = \varphi(w) + \varphi(z)$$

for all $w, z \in \mathbf{C}$ but φ is not linear. (Here $\mathbf{C}$ is thought of as a complex vector space.)
[*There also exists a function $\varphi\colon \mathbf{R} \to \mathbf{R}$ such that φ satisfies the additivity condition above but φ is not linear. However, showing the existence of such a function involves considerably more advanced tools.*]

10 Suppose U is a subspace of V with $U \neq V$. Suppose $S \in \mathcal{L}(U, W)$ and $S \neq 0$ (which means that $Su \neq 0$ for some $u \in U$). Define $T\colon V \to W$ by

$$Tv = \begin{cases} Sv & \text{if } v \in U, \\ 0 & \text{if } v \in V \text{ and } v \notin U. \end{cases}$$

Prove that T is not a linear map on V.

11 Suppose V is finite-dimensional. Prove that every linear map on a subspace of V can be extended to a linear map on V. In other words, show that if U is a subspace of V and $S \in \mathcal{L}(U, W)$, then there exists $T \in \mathcal{L}(V, W)$ such that $Tu = Su$ for all $u \in U$.

12 Suppose V is finite-dimensional with $\dim V > 0$, and suppose W is infinite-dimensional. Prove that $\mathcal{L}(V, W)$ is infinite-dimensional.

13 Suppose $v_1, \dots, v_m$ is a linearly dependent list of vectors in V. Suppose also that $W \neq \{0\}$. Prove that there exist $w_1, \dots, w_m \in W$ such that no $T \in \mathcal{L}(V, W)$ satisfies $Tv_k = w_k$ for each $k = 1, \dots, m$.

14 Suppose V is finite-dimensional with $\dim V \geq 2$. Prove that there exist $S, T \in \mathcal{L}(V, V)$ such that $ST \neq TS$.

3.B *Null Spaces and Ranges*

Null Space and Injectivity

In this section we will learn about two subspaces that are intimately connected with each linear map. We begin with the set of vectors that get mapped to 0.

3.12 Definition ***null space,*** null T

For $T \in \mathcal{L}(V, W)$, the ***null space*** of T, denoted null T, is the subset of V consisting of those vectors that T maps to 0:

$$\operatorname{null} T = \{v \in V : Tv = 0\}.$$

3.13 Example ***null space***

- If T is the zero map from V to W, in other words if $Tv = 0$ for every $v \in V$, then null $T = V$.
- Suppose $\varphi \in \mathcal{L}(\mathbf{C}^3, \mathbf{F})$ is defined by $\varphi(z_1, z_2, z_3) = z_1 + 2z_2 + 3z_3$. Then $\operatorname{null} \varphi = \{(z_1, z_2, z_3) \in \mathbf{C}^3 : z_1 + 2z_2 + 3z_3 = 0\}$. A basis of null φ is $(-2, 1, 0), (-3, 0, 1)$.
- Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. The only functions whose derivative equals the zero function are the constant functions. Thus the null space of D equals the set of constant functions.
- Suppose $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the multiplication by x^2 map defined by $(Tp)(x) = x^2 p(x)$. The only polynomial p such that $x^2 p(x) = 0$ for all $x \in \mathbf{R}$ is the 0 polynomial. Thus null $T = \{0\}$.
- Suppose $T \in \mathcal{L}(\mathbf{F}^\infty, \mathbf{F}^\infty)$ is the backward shift defined by

  $$T(x_1, x_2, x_3, \dots) = (x_2, x_3, \dots).$$

  Clearly $T(x_1, x_2, x_3, \dots)$ equals 0 if and only if $x_2, x_3, \dots$ are all 0. Thus in this case we have $\operatorname{null} T = \{(a, 0, 0, \dots) : a \in \mathbf{F}\}$.

The next result shows that the null space of each linear map is a subspace of the domain. In particular, 0 is in the null space of every linear map.

*Some mathematicians use the term **kernel** instead of null space. The word "null" means zero. Thus the term "null space" should remind you of the connection to* 0.

3.14 The null space is a subspace

Suppose $T \in \mathcal{L}(V, W)$. Then null T is a subspace of V.

Proof Because T is a linear map, we know that $T(0) = 0$ (by 3.11). Thus $0 \in \operatorname{null} T$.

Suppose $u, v \in \operatorname{null} T$. Then

$$T(u + v) = Tu + Tv = 0 + 0 = 0.$$

Hence $u + v \in \operatorname{null} T$. Thus null T is closed under addition.

Suppose $u \in \operatorname{null} T$ and $\lambda \in \mathbf{F}$. Then

$$T(\lambda u) = \lambda Tu = \lambda 0 = 0.$$

Hence $\lambda u \in \operatorname{null} T$. Thus null T is closed under scalar multiplication.

Take another look at the null spaces that were computed in Example 3.13 and note that all of them are subspaces.

We have shown that null T contains 0 and is closed under addition and scalar multiplication. Thus null T is a subspace of V (by 1.34). ■

As we will soon see, for a linear map the next definition is closely connected to the null space.

3.15 Definition ***injective***

A function $T: V \to W$ is called ***injective*** if $Tu = Tv$ implies $u = v$.

*Many mathematicians use the term **one-to-one**, which means the same as injective.*

The definition above could be rephrased to say that T is injective if $u \neq v$ implies that $Tu \neq Tv$. In other words, T is injective if it maps distinct inputs to distinct outputs.

The next result says that we can check whether a linear map is injective by checking whether 0 is the only vector that gets mapped to 0. As a simple application of this result, we see that of the linear maps whose null spaces we computed in 3.13, only multiplication by x^2 is injective (except that the zero map is injective in the special case $V = \{0\}$).

3.16 Injectivity is equivalent to null space equals $\{0\}$

Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if $\operatorname{null} T = \{0\}$.

Proof First suppose T is injective. We want to prove that $\operatorname{null} T = \{0\}$. We already know that $\{0\} \subset \operatorname{null} T$ (by 3.11). To prove the inclusion in the other direction, suppose $v \in \operatorname{null} T$. Then

$$T(v) = 0 = T(0).$$

Because T is injective, the equation above implies that $v = 0$. Thus we can conclude that $\operatorname{null} T = \{0\}$, as desired.

To prove the implication in the other direction, now suppose $\operatorname{null} T = \{0\}$. We want to prove that T is injective. To do this, suppose $u, v \in V$ and $Tu = Tv$. Then

$$0 = Tu - Tv = T(u - v).$$

Thus $u - v$ is in $\operatorname{null} T$, which equals $\{0\}$. Hence $u - v = 0$, which implies that $u = v$. Hence T is injective, as desired. ∎

Range and Surjectivity

Now we give a name to the set of outputs of a function.

3.17 Definition ***range***

For T a function from V to W, the ***range*** of T is the subset of W consisting of those vectors that are of the form Tv for some $v \in V$:

$$\operatorname{range} T = \{Tv : v \in V\}.$$

3.18 Example ***range***

- If T is the zero map from V to W, in other words if $Tv = 0$ for every $v \in V$, then $\operatorname{range} T = \{0\}$.

- Suppose $T \in \mathcal{L}(\mathbf{R}^2, \mathbf{R}^3)$ is defined by $T(x, y) = (2x, 5y, x + y)$, then $\operatorname{range} T = \{(2x, 5y, x + y) : x, y \in \mathbf{R}\}$. A basis of $\operatorname{range} T$ is $(2, 0, 1), (0, 5, 1)$.

- Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. Because for every polynomial $q \in \mathcal{P}(\mathbf{R})$ there exists a polynomial $p \in \mathcal{P}(\mathbf{R})$ such that $p' = q$, the range of D is $\mathcal{P}(\mathbf{R})$.

*Some mathematicians use the word **image**, which means the same as range.*

The next result shows that the range of each linear map is a subspace of the vector space into which it is being mapped.

3.19 The range is a subspace

If $T \in \mathcal{L}(V, W)$, then range T is a subspace of W.

Proof Suppose $T \in \mathcal{L}(V, W)$. Then $T(0) = 0$ (by 3.11), which implies that $0 \in \operatorname{range} T$.

If $w_1, w_2 \in \operatorname{range} T$, then there exist $v_1, v_2 \in V$ such that $Tv_1 = w_1$ and $Tv_2 = w_2$. Thus

$$T(v_1 + v_2) = Tv_1 + Tv_2 = w_1 + w_2.$$

Hence $w_1 + w_2 \in \operatorname{range} T$. Thus range T is closed under addition.

If $w \in \operatorname{range} T$ and $\lambda \in \mathbf{F}$, then there exists $v \in V$ such that $Tv = w$. Thus

$$T(\lambda v) = \lambda T v = \lambda w.$$

Hence $\lambda w \in \operatorname{range} T$. Thus range T is closed under scalar multiplication.

We have shown that range T contains 0 and is closed under addition and scalar multiplication. Thus range T is a subspace of W (by 1.34). ■

3.20 Definition ***surjective***

A function $T: V \to W$ is called ***surjective*** if its range equals W.

To illustrate the definition above, note that of the ranges we computed in 3.18, only the differentiation map is surjective (except that the zero map is surjective in the special case $W = \{0\}$.

*Many mathematicians use the term **onto**, which means the same as surjective.*

Whether a linear map is surjective depends on what we are thinking of as the vector space into which it maps.

3.21 Example The differentiation map $D \in \mathcal{L}\big(\mathcal{P}_5(\mathbf{R}), \mathcal{P}_5(\mathbf{R})\big)$ defined by $Dp = p'$ is not surjective, because the polynomial x^5 is not in the range of D. However, the differentiation map $S \in \mathcal{L}\big(\mathcal{P}_5(\mathbf{R}), \mathcal{P}_4(\mathbf{R})\big)$ defined by $Sp = p'$ is surjective, because its range equals $\mathcal{P}_4(\mathbf{R})$, which is now the vector space into which S maps.

Fundamental Theorem of Linear Maps

The next result is so important that it gets a dramatic name.

3.22 Fundamental Theorem of Linear Maps

Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then range T is finite-dimensional and

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

Proof Let $u_1, \ldots, u_m$ be a basis of null T; thus $\dim \operatorname{null} T = m$. The linearly independent list $u_1, \ldots, u_m$ can be extended to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_n$$

of V (by 2.33). Thus $\dim V = m + n$. To complete the proof, we need only show that range T is finite-dimensional and $\dim \operatorname{range} T = n$. We will do this by proving that $Tv_1, \ldots, Tv_n$ is a basis of range T.

Let $v \in V$. Because $u_1, \ldots, u_m, v_1, \ldots, v_n$ spans V, we can write

$$v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n,$$

where the a's and b's are in $\mathbf{F}$. Applying T to both sides of this equation, we get

$$Tv = b_1Tv_1 + \cdots + b_nTv_n,$$

where the terms of the form Tu_j disappeared because each u_j is in null T. The last equation implies that $Tv_1, \ldots, Tv_n$ spans range T. In particular, range T is finite-dimensional.

To show $Tv_1, \ldots, Tv_n$ is linearly independent, suppose $c_1, \ldots, c_n \in \mathbf{F}$ and

$$c_1Tv_1 + \cdots + c_nTv_n = 0.$$

Then

$$T(c_1v_1 + \cdots + c_nv_n) = 0.$$

Hence

$$c_1v_1 + \cdots + c_nv_n \in \operatorname{null} T.$$

Because $u_1, \ldots, u_m$ spans null T, we can write

$$c_1v_1 + \cdots + c_nv_n = d_1u_1 + \cdots + d_mu_m,$$

where the d's are in $\mathbf{F}$. This equation implies that all the c's (and d's) are 0 (because $u_1, \ldots, u_m, v_1, \ldots, v_n$ is linearly independent). Thus $Tv_1, \ldots, Tv_n$ is linearly independent and hence is a basis of range T, as desired. ■

Now we can show that no linear map from a finite-dimensional vector space to a "smaller" vector space can be injective, where "smaller" is measured by dimension.

3.23 A map to a smaller dimensional space is not injective

Suppose V and W are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from V to W is injective.

Proof Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &\geq \dim V - \dim W \\ &> 0, \end{aligned}$$

where the equality above comes from the Fundamental Theorem of Linear Maps (3.22). The inequality above states that $\dim \operatorname{null} T > 0$. This means that $\operatorname{null} T$ contains vectors other than 0. Thus T is not injective (by 3.16). ■

The next result shows that no linear map from a finite-dimensional vector space to a "bigger" vector space can be surjective, where "bigger" is measured by dimension.

3.24 A map to a larger dimensional space is not surjective

Suppose V and W are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from V to W is surjective.

Proof Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T \\ &\leq \dim V \\ &< \dim W, \end{aligned}$$

where the equality above comes from the Fundamental Theorem of Linear Maps (3.22). The inequality above states that $\dim \operatorname{range} T < \dim W$. This means that $\operatorname{range} T$ cannot equal W. Thus T is not surjective. ■

As we will now see, 3.23 and 3.24 have important consequences in the theory of linear equations. The idea here is to express questions about systems of linear equations in terms of linear maps.

3.25 Example Rephrase in terms of a linear map the question of whether a homogeneous system of linear equations has a nonzero solution.

Solution

Homogeneous, in this context, means that the constant term on the right side of each equation below is 0.

Fix positive integers m and n, and let $A_{j,k} \in \mathbf{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$. Consider the homogeneous system of linear equations

$$\sum_{k=1}^{n} A_{1,k} x_k = 0$$
$$\vdots$$
$$\sum_{k=1}^{n} A_{m,k} x_k = 0.$$

Obviously $x_1 = \cdots = x_n = 0$ is a solution of the system of equations above; the question here is whether any other solutions exist.

Define $T \colon \mathbf{F}^n \to \mathbf{F}^m$ by

$$T(x_1, \dots, x_n) = \Big(\sum_{k=1}^{n} A_{1,k} x_k, \dots, \sum_{k=1}^{n} A_{m,k} x_k\Big).$$

The equation $T(x_1, \dots, x_n) = 0$ (the 0 here is the additive identity in $\mathbf{F}^m$, namely, the list of length m of all 0's) is the same as the homogeneous system of linear equations above.

Thus we want to know if null T is strictly bigger than $\{0\}$. In other words, we can rephrase our question about nonzero solutions as follows (by 3.16): What condition ensures that T is not injective?

3.26 Homogeneous system of linear equations

A homogeneous system of linear equations with more variables than equations has nonzero solutions.

Proof Use the notation and result from the example above. Thus T is a linear map from $\mathbf{F}^n$ to $\mathbf{F}^m$, and we have a homogeneous system of m linear equations with n variables $x_1, \dots, x_n$. From 3.23 we see that T is not injective if $n > m$. ■

Example of the result above: a homogeneous system of four linear equations with five variables has nonzero solutions.

3.27 Example Consider the question of whether an inhomogeneous system of linear equations has no solutions for some choice of the constant terms. Rephrase this question in terms of a linear map.

Solution Fix positive integers m and n, and let $A_{j,k} \in \mathbf{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$. For $c_1, \dots, c_m \in \mathbf{F}$, consider the system of linear equations

3.28
$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= c_1 \\ &\vdots \\ \sum_{k=1}^{n} A_{m,k} x_k &= c_m. \end{aligned}$$

The question here is whether there is some choice of $c_1, \dots, c_m \in \mathbf{F}$ such that no solution exists to the system above.

Define $T : \mathbf{F}^n \to \mathbf{F}^m$ by

$$T(x_1, \dots, x_n) = \Big(\sum_{k=1}^{n} A_{1,k} x_k, \dots, \sum_{k=1}^{n} A_{m,k} x_k\Big).$$

The equation $T(x_1, \dots, x_n) = (c_1, \dots, c_m)$ is the same as the system of equations 3.28. Thus we want to know if range $T \neq \mathbf{F}^m$. Hence we can rephrase our question about not having a solution for some choice of $c_1, \dots, c_m \in \mathbf{F}$ as follows: What condition ensures that T is not surjective?

3.29 Inhomogeneous system of linear equations

An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.

Our results about homogeneous systems with more variables than equations and inhomogeneous systems with more equations than variables (3.26 and 3.29) are often proved using Gaussian elimination. The abstract approach taken here leads to cleaner proofs.

Proof Use the notation and result from the example above. Thus T is a linear map from $\mathbf{F}^n$ to $\mathbf{F}^m$, and we have a system of m equations with n variables $x_1, \dots, x_n$. From 3.24 we see that T is not surjective if $n < m$. ∎

Example of the result above: an inhomogeneous system of five linear equations with four variables has no solution for some choice of the constant terms.

EXERCISES 3.B

1 Give an example of a linear map T such that $\dim \operatorname{null} T = 3$ and $\dim \operatorname{range} T = 2$.

2 Suppose V is a vector space and $S, T \in \mathcal{L}(V, V)$ are such that

$$\operatorname{range} S \subset \operatorname{null} T.$$

Prove that $(ST)^2 = 0$.

3 Suppose $v_1, \ldots, v_m$ is a list of vectors in V. Define $T \in \mathcal{L}(\mathbf{F}^m, V)$ by

$$T(z_1, \ldots, z_m) = z_1 v_1 + \cdots + z_m v_m.$$

(a) What property of T corresponds to $v_1, \ldots, v_m$ spanning V?

(b) What property of T corresponds to $v_1, \ldots, v_m$ being linearly independent?

4 Show that

$$\{T \in \mathcal{L}(\mathbf{R}^5, \mathbf{R}^4) : \dim \operatorname{null} T > 2\}$$

is not a subspace of $\mathcal{L}(\mathbf{R}^5, \mathbf{R}^4)$.

5 Give an example of a linear map $T : \mathbf{R}^4 \to \mathbf{R}^4$ such that

$$\operatorname{range} T = \operatorname{null} T.$$

(x, y, z, t) -> (0, x, 0, z)

6 Prove that there does not exist a linear map $T : \mathbf{R}^5 \to \mathbf{R}^5$ such that

$$\operatorname{range} T = \operatorname{null} T.$$

7 Suppose V and W are finite-dimensional with $2 \leq \dim V \leq \dim W$. Show that $\{T \in \mathcal{L}(V, W) : T \text{ is not injective}\}$ is not a subspace of $\mathcal{L}(V, W)$.

8 Suppose V and W are finite-dimensional with $\dim V \geq \dim W \geq 2$. Show that $\{T \in \mathcal{L}(V, W) : T \text{ is not surjective}\}$ is not a subspace of $\mathcal{L}(V, W)$.

9 Suppose $T \in \mathcal{L}(V, W)$ is injective and $v_1, \ldots, v_n$ is linearly independent in V. Prove that $Tv_1, \ldots, Tv_n$ is linearly independent in W.

10 Suppose $v_1, \ldots, v_n$ spans V and $T \in \mathcal{L}(V, W)$. Prove that the list $Tv_1, \ldots, Tv_n$ spans range T.

11 Suppose $S_1, \ldots, S_n$ are injective linear maps such that $S_1 S_2 \cdots S_n$ makes sense. Prove that $S_1 S_2 \cdots S_n$ is injective.

12 Suppose that V is finite-dimensional and that $T \in \mathcal{L}(V, W)$. Prove that there exists a subspace U of V such that $U \cap \operatorname{null} T = \{0\}$ and $\operatorname{range} T = \{Tu : u \in U\}$.

13 Suppose T is a linear map from $\mathbf{F}^4$ to $\mathbf{F}^2$ such that

$$\operatorname{null} T = \{(x_1, x_2, x_3, x_4) \in \mathbf{F}^4 : x_1 = 5x_2 \text{ and } x_3 = 7x_4\}.$$

Prove that T is surjective.

14 Suppose U is a 3-dimensional subspace of $\mathbf{R}^8$ and that T is a linear map from $\mathbf{R}^8$ to $\mathbf{R}^5$ such that $\operatorname{null} T = U$. Prove that T is surjective.

15 Prove that there does not exist a linear map from $\mathbf{F}^5$ to $\mathbf{F}^2$ whose null space equals

$$\{(x_1, x_2, x_3, x_4, x_5) \in \mathbf{F}^5 : x_1 = 3x_2 \text{ and } x_3 = x_4 = x_5\}.$$

16 Suppose there exists a linear map on V whose null space and range are both finite-dimensional. Prove that V is finite-dimensional.

17 Suppose V and W are both finite-dimensional. Prove that there exists an injective linear map from V to W if and only if $\dim V \leq \dim W$.

18 Suppose V and W are both finite-dimensional. Prove that there exists a surjective linear map from V onto W if and only if $\dim V \geq \dim W$.

19 Suppose V and W are finite-dimensional and that U is a subspace of V. Prove that there exists $T \in \mathcal{L}(V, W)$ such that $\operatorname{null} T = U$ if and only if $\dim U \geq \dim V - \dim W$.

20 Suppose W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is injective if and only if there exists $S \in \mathcal{L}(W, V)$ such that ST is the identity map on V.

21 Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that T is surjective if and only if there exists $S \in \mathcal{L}(W, V)$ such that TS is the identity map on W.

22 Suppose U and V are finite-dimensional vector spaces and $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(U, V)$. Prove that

$$\dim \operatorname{null} ST \le \dim \operatorname{null} S + \dim \operatorname{null} T.$$

23 Suppose U and V are finite-dimensional vector spaces and $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(U, V)$. Prove that

$$\dim \operatorname{range} ST \le \min\{\dim \operatorname{range} S, \dim \operatorname{range} T\}.$$

24 Suppose W is finite-dimensional and $T_1, T_2 \in \mathcal{L}(V, W)$. Prove that $\operatorname{null} T_1 \subset \operatorname{null} T_2$ if and only if there exists $S \in \mathcal{L}(W, W)$ such that $T_2 = ST_1$.

25 Suppose V is finite-dimensional and $T_1, T_2 \in \mathcal{L}(V, W)$. Prove that $\operatorname{range} T_1 \subset \operatorname{range} T_2$ if and only if there exists $S \in \mathcal{L}(V, V)$ such that $T_1 = T_2 S$.

26 Suppose $D \in \mathcal{L}\big(\mathcal{P}(\mathbf{R}), \mathcal{P}(\mathbf{R})\big)$ is such that $\deg Dp = (\deg p) - 1$ for every nonconstant polynomial $p \in \mathcal{P}(\mathbf{R})$. Prove that D is surjective.
[*The notation D is used above to remind you of the differentiation map that sends a polynomial p to p'. Without knowing the formula for the derivative of a polynomial (except that it reduces the degree by 1), you can use the exercise above to show that for every polynomial $q \in \mathcal{P}(\mathbf{R})$, there exists a polynomial $p \in \mathcal{P}(\mathbf{R})$ such that $p' = q$.*]

27 Suppose $p \in \mathcal{P}(\mathbf{R})$. Prove that there exists a polynomial $q \in \mathcal{P}(\mathbf{R})$ such that $5q'' + 3q' = p$.
[*This exercise can be done without linear algebra, but it's more fun to do it using linear algebra.*]

28 Suppose $T \in \mathcal{L}(V, W)$, and $w_1, \dots, w_m$ is a basis of $\operatorname{range} T$. Prove that there exist $\varphi_1, \dots, \varphi_m \in \mathcal{L}(V, \mathbf{F})$ such that

$$Tv = \varphi_1(v)w_1 + \cdots + \varphi_m(v)w_m$$

for every $v \in V$.

29 Suppose $\varphi \in \mathcal{L}(V, \mathbf{F})$. Suppose $u \in V$ is not in $\operatorname{null} \varphi$. Prove that

$$V = \operatorname{null} \varphi \oplus \{au : a \in \mathbf{F}\}.$$

30 Suppose φ_1 and φ_2 are linear maps from V to $\mathbf{F}$ that have the same null space. Show that there exists a constant $c \in \mathbf{F}$ such that $\varphi_1 = c\varphi_2$.

31 Give an example of two linear maps T_1 and T_2 from $\mathbf{R}^5$ to $\mathbf{R}^2$ that have the same null space but are such that T_1 is not a scalar multiple of T_2.

3.C Matrices

Representing a Linear Map by a Matrix

We know that if $v_1, \dots, v_n$ is a basis of V and $T : V \to W$ is linear, then the values of $Tv_1, \dots, Tv_n$ determine the values of T on arbitrary vectors in V (see 3.5). As we will soon see, matrices are used as an efficient method of recording the values of the Tv_j's in terms of a basis of W.

3.30 Definition ***matrix,*** $A_{j,k}$

Let m and n denote positive integers. An m-by-n ***matrix*** A is a rectangular array of elements of $\mathbf{F}$ with m rows and n columns:

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix}.$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A. In other words, the first index refers to the row number and the second index refers to the column number.

Thus $A_{2,3}$ refers to the entry in the second row, third column of a matrix A.

3.31 Example If $A = \begin{pmatrix} 8 & 4 & 5-3i \\ 1 & 9 & 7 \end{pmatrix}$, then $A_{2,3} = 7$.

Now we come to the key definition in this section.

3.32 Definition ***matrix of a linear map,*** $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. The ***matrix of*** T with respect to these bases is the m-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases are not clear from the context, then the notation $\mathcal{M}\big(T, (v_1, \dots, v_n), (w_1, \dots, w_m)\big)$ is used.

The matrix $\mathcal{M}(T)$ of a linear map $T \in \mathcal{L}(V, W)$ depends on the basis $v_1, \dots, v_n$ of V and the basis $w_1, \dots, w_m$ of W, as well as on T. However, the bases should be clear from the context, and thus they are often not included in the notation.

To remember how $\mathcal{M}(T)$ is constructed from T, you might write across the top of the matrix the basis vectors $v_1, \dots, v_n$ for the domain and along the left the basis vectors $w_1, \dots, w_m$ for the vector space into which T maps, as follows:

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \!\!\! \begin{array}{c} \begin{array}{ccccc} v_1 & \dots & v_k & \dots & v_n \end{array} \\ \left(\begin{array}{ccccc} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{array} \right) \end{array}.$$

In the matrix above only the k^{th} column is shown. Thus the second index of each displayed entry of the matrix above is k. The picture above should remind you that Tv_k can be computed from $\mathcal{M}(T)$ by multiplying each entry in the k^{th} column by the corresponding w_j from the left column, and then adding up the resulting vectors.

The k^{th} column of $\mathcal{M}(T)$ consists of the scalars needed to write Tv_k as a linear combination of $(w_1, \dots, w_m)$:

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j.$$

If T is a linear map from $\mathbf{F}^n$ to $\mathbf{F}^m$, then unless stated otherwise, assume the bases in question are the standard ones (where the k^{th} basis vector is 1 in the k^{th} slot and 0 in all the other slots). If you think of elements of $\mathbf{F}^m$ as columns of m numbers, then you can think of the k^{th} column of $\mathcal{M}(T)$ as T applied to the k^{th} standard basis vector.

If T maps an n-dimensional vector space to an m-dimensional vector space, then $\mathcal{M}(T)$ is an m-by-n matrix.

3.33 Example Suppose $T \in \mathcal{L}(\mathbf{F}^2, \mathbf{F}^3)$ is defined by

$$T(x, y) = (x + 3y, 2x + 5y, 7x + 9y).$$

Find the matrix of T with respect to the standard bases of $\mathbf{F}^2$ and $\mathbf{F}^3$.

Solution Because $T(1,0) = (1,2,7)$ and $T(0,1) = (3,5,9)$, the matrix of T with respect to the standard bases is the 3-by-2 matrix below:

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}.$$

When working with $\mathcal{P}_m(\mathbf{F})$, use the standard basis $1, x, x^2, \ldots, x^m$ unless the context indicates otherwise.

3.34 **Example** Suppose $D \in \mathcal{L}\big(\mathcal{P}_3(\mathbf{R}), \mathcal{P}_2(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. Find the matrix of D with respect to the standard bases of $\mathcal{P}_3(\mathbf{R})$ and $\mathcal{P}_2(\mathbf{R})$.

Solution Because $(x^n)' = nx^{n-1}$, the matrix of T with respect to the standard bases is the 3-by-4 matrix below:

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Addition and Scalar Multiplication of Matrices

For the rest of this section, assume that V and W are finite-dimensional and that a basis has been chosen for each of these vector spaces. Thus for each linear map from V to W, we can talk about its matrix (with respect to the chosen bases, of course). Is the matrix of the sum of two linear maps equal to the sum of the matrices of the two maps?

Right now this question does not make sense, because although we have defined the sum of two linear maps, we have not defined the sum of two matrices. Fortunately, the obvious definition of the sum of two matrices has the right properties. Specifically, we make the following definition.

3.35 **Definition** ***matrix addition***

The ***sum of two matrices of the same size*** is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} A_{1,1} & \ldots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \ldots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \ldots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \ldots & C_{m,n} \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1} + C_{1,1} & \ldots & A_{1,n} + C_{1,n} \\ \vdots & & \vdots \\ A_{m,1} + C_{m,1} & \ldots & A_{m,n} + C_{m,n} \end{pmatrix}.$$

In other words, $(A + C)_{j,k} = A_{j,k} + C_{j,k}$.

In the following result, the assumption is that the same bases are used for all three linear maps $S + T$, S, and T.

3.36 The matrix of the sum of linear maps

Suppose $S, T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$.

The verification of the result above is left to the reader.

Still assuming that we have some bases in mind, is the matrix of a scalar times a linear map equal to the scalar times the matrix of the linear map? Again the question does not make sense, because we have not defined scalar multiplication on matrices. Fortunately, the obvious definition again has the right properties.

3.37 Definition ***scalar multiplication of a matrix***

The product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \dots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \dots & \lambda A_{m,n} \end{pmatrix}.$$

In other words, $(\lambda A)_{j,k} = \lambda A_{j,k}$.

In the following result, the assumption is that the same bases are used for both linear maps λT and T.

3.38 The matrix of a scalar times a linear map

Suppose $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

The verification of the result above is also left to the reader.

Because addition and scalar multiplication have now been defined for matrices, you should not be surprised that a vector space is about to appear. We need only a bit of notation so that this new vector space has a name.

3.39 Notation $\mathbf{F}^{m,n}$

For m and n positive integers, the set of all m-by-n matrices with entries in $\mathbf{F}$ is denoted by $\mathbf{F}^{m,n}$.

3.40 $\dim \mathbf{F}^{m,n} = mn$

Suppose m and n are positive integers. With addition and scalar multiplication defined as above, $\mathbf{F}^{m,n}$ is a vector space with dimension mn.

Proof The verification that $\mathbf{F}^{m,n}$ is a vector space is left to the reader. Note that the additive identity of $\mathbf{F}^{m,n}$ is the m-by-n matrix whose entries all equal 0.

The reader should also verify that the list of m-by-n matrices that have 0 in all entries except for a 1 in one entry is a basis of $\mathbf{F}^{m,n}$. There are mn such matrices, so the dimension of $\mathbf{F}^{m,n}$ equals mn. ■

Matrix Multiplication

Suppose, as previously, that $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. Suppose also that we have another vector space U and that $u_1, \dots, u_p$ is a basis of U.

Consider linear maps $T: U \to V$ and $S: V \to W$. The composition ST is a linear map from U to W. Does $\mathcal{M}(ST)$ equal $\mathcal{M}(S)\mathcal{M}(T)$? This question does not yet make sense, because we have not defined the product of two matrices. We will choose a definition of matrix multiplication that forces this question to have a positive answer. Let's see how to do this.

Suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$. For $1 \le k \le p$, we have

$$\begin{aligned}
(ST)u_k &= S\Big(\sum_{r=1}^{n} C_{r,k} v_r\Big) \\
&= \sum_{r=1}^{n} C_{r,k} S v_r \\
&= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r} w_j \\
&= \sum_{j=1}^{m} \Big(\sum_{r=1}^{n} A_{j,r} C_{r,k}\Big) w_j .
\end{aligned}$$

Thus $\mathcal{M}(ST)$ is the m-by-p matrix whose entry in row j, column k, equals

$$\sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Now we see how to define matrix multiplication so that the desired equation $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$ holds.

3.41 Definition ***matrix multiplication***

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then AC is defined to be the m-by-p matrix whose entry in row j, column k, is given by the following equation:

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

In other words, the entry in row j, column k, of AC is computed by taking row j of A and column k of C, multiplying together corresponding entries, and then summing.

Note that we define the product of two matrices only when the number of columns of the first matrix equals the number of rows of the second matrix.

You may have learned this definition of matrix multiplication in an earlier course, although you may not have seen the motivation for it.

3.42 Example Here we multiply together a 3-by-2 matrix and a 2-by-4 matrix, obtaining a 3-by-4 matrix:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{pmatrix}.$$

Matrix multiplication is not commutative. In other words, AC is not necessarily equal to CA even if both products are defined (see Exercise 12). Matrix multiplication is distributive and associative (see Exercises 13 and 14).

In the following result, the assumption is that the same basis of V is used in considering $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, the same basis of W is used in considering $S \in \mathcal{L}(V, W)$ and $ST \in \mathcal{L}(U, W)$, and the same basis of U is used in considering $T \in \mathcal{L}(U, V)$ and $ST \in \mathcal{L}(U, W)$.

3.43 The matrix of the product of linear maps

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.

The proof of the result above is the calculation that was done as motivation before the definition of matrix multiplication.

In the next piece of notation, note that as usual the first index refers to a row and the second index refers to a column, with a vertically centered dot used as a placeholder.

3.44 Notation $A_{j,\cdot}$, $A_{\cdot,k}$

Suppose A is an m-by-n matrix.

- If $1 \leq j \leq m$, then $A_{j,\cdot}$ denotes the 1-by-n matrix consisting of row j of A.
- If $1 \leq k \leq n$, then $A_{\cdot,k}$ denotes the m-by-1 matrix consisting of column k of A.

3.45 Example If $A = \begin{pmatrix} 8 & 4 & 5 \\ 1 & 9 & 7 \end{pmatrix}$, then $A_{2,\cdot}$ is row 2 of A and $A_{\cdot,2}$ is column 2 of A. In other words,

$$A_{2,\cdot} = \begin{pmatrix} 1 & 9 & 7 \end{pmatrix} \quad \text{and} \quad A_{\cdot,2} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}.$$

The product of a 1-by-n matrix and an n-by-1 matrix is a 1-by-1 matrix. However, we will frequently identify a 1-by-1 matrix with its entry.

3.46 Example $\begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 26 \end{pmatrix}$ because $3 \cdot 6 + 4 \cdot 2 = 26$.

However, we can identify $\begin{pmatrix} 26 \end{pmatrix}$ with 26, writing $\begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = 26$.

Our next result gives another way to think of matrix multiplication: the entry in row j, column k, of AC equals (row j of A) times (column k of C).

3.47 Entry of matrix product equals row times column

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{j,k} = A_{j,\cdot} C_{\cdot,k}$$

for $1 \leq j \leq m$ and $1 \leq k \leq p$.

The proof of the result above follows immediately from the definitions.

3.48 Example The result above and Example 3.46 show why the entry in row 2, column 1, of the product in Example 3.42 equals 26.

The next result gives yet another way to think of matrix multiplication. It states that column k of AC equals A times column k of C.

3.49 Column of matrix product equals matrix times column

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{\cdot,k} = AC_{\cdot,k}$$

for $1 \leq k \leq p$.

Again, the proof of the result above follows immediately from the definitions and is left to the reader.

3.50 **Example** From the result above and the equation

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 19 \\ 31 \end{pmatrix},$$

we see why column 2 in the matrix product in Example 3.42 is the right side of the equation above.

We give one more way of thinking about the product of an m-by-n matrix and an n-by-1 matrix. The following example illustrates this approach.

3.51 **Example** In the example above, the product of a 3-by-2 matrix and a 2-by-1 matrix is a linear combination of the columns of the 3-by-2 matrix, with the scalars that multiply the columns coming from the 2-by-1 matrix. Specifically,

$$\begin{pmatrix} 7 \\ 19 \\ 31 \end{pmatrix} = 5\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + 1\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}.$$

The next result generalizes the example above. Again, the proof follows easily from the definitions and is left to the reader.

3.52 Linear combination of columns

Suppose A is an m-by-n matrix and $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ is an n-by-1 matrix.

Then

$$Ac = c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}.$$

In other words, Ac is a linear combination of the columns of A, with the scalars that multiply the columns coming from c.

Two more ways to think about matrix multiplication are given by Exercises 10 and 11.

EXERCISES 3.C

1 Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Show that with respect to each choice of bases of V and W, the matrix of T has at least $\dim \operatorname{range} T$ nonzero entries.

2 Suppose $D \in \mathcal{L}\big(\mathcal{P}_3(\mathbf{R}), \mathcal{P}_2(\mathbf{R})\big)$ is the differentiation map defined by $Dp = p'$. Find a basis of $\mathcal{P}_3(\mathbf{R})$ and a basis of $\mathcal{P}_2(\mathbf{R})$ such that the matrix of D with respect to these bases is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

[*Compare the exercise above to Example 3.34. The next exercise generalizes the exercise above.*]

3 Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that there exist a basis of V and a basis of W such that with respect to these bases, all entries of $\mathcal{M}(T)$ are 0 except that the entries in row j, column j, equal 1 for $1 \le j \le \dim \operatorname{range} T$.

4 Suppose $v_1, \dots, v_m$ is a basis of V and W is finite-dimensional. Suppose $T \in \mathcal{L}(V, W)$. Prove that there exists a basis $w_1, \dots, w_n$ of W such that all the entries in the first column of $\mathcal{M}(T)$ (with respect to the bases $v_1, \dots, v_m$ and $w_1, \dots, w_n$) are 0 except for possibly a 1 in the first row, first column.
[*In this exercise, unlike Exercise 3, you are given the basis of V instead of being able to choose a basis of V.*]

5 Suppose $w_1, \dots, w_n$ is a basis of W and V is finite-dimensional. Suppose $T \in \mathcal{L}(V, W)$. Prove that there exists a basis $v_1, \dots, v_m$ of V such that all the entries in the first row of $\mathcal{M}(T)$ (with respect to the bases $v_1, \dots, v_m$ and $w_1, \dots, w_n$) are 0 except for possibly a 1 in the first row, first column.
[*In this exercise, unlike Exercise 3, you are given the basis of W instead of being able to choose a basis of W.*]

6 Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that $\dim \operatorname{range} T = 1$ if and only if there exist a basis of V and a basis of W such that with respect to these bases, all entries of $\mathcal{M}(T)$ equal 1.

7 Verify 3.36.

8 Verify 3.38.

9 Prove 3.52.

10 Suppose A is an m-by-n matrix and C is an n-by-p matrix. Prove that

$$(AC)_{j,\cdot} = A_{j,\cdot}C$$

for $1 \leq j \leq m$. In other words, show that row j of AC equals (row j of A) times C.

11 Suppose $a = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$ is a 1-by-n matrix and C is an n-by-p matrix. Prove that

$$aC = a_1 C_{1,\cdot} + \cdots + a_n C_{n,\cdot}.$$

In other words, show that aC is a linear combination of the rows of C, with the scalars that multiply the rows coming from a.

12 Give an example with 2-by-2 matrices to show that matrix multiplication is not commutative. In other words, find 2-by-2 matrices A and C such that $AC \neq CA$.

13 Prove that the distributive property holds for matrix addition and matrix multiplication. In other words, suppose A, B, C, D, E, and F are matrices whose sizes are such that $A(B + C)$ and $(D + E)F$ make sense. Prove that $AB + AC$ and $DF + EF$ both make sense and that $A(B + C) = AB + AC$ and $(D + E)F = DF + EF$.

14 Prove that matrix multiplication is associative. In other words, suppose A, B, and C are matrices whose sizes are such that $(AB)C$ makes sense. Prove that $A(BC)$ makes sense and that $(AB)C = A(BC)$.

15 Suppose A is an n-by-n matrix and $1 \leq j, k \leq n$. Show that the entry in row j, column k, of A^3 (which is defined to mean AAA) is

$$\sum_{p=1}^{n} \sum_{r=1}^{n} A_{j,p} A_{p,r} A_{r,k}.$$

3.D *Invertibility and Isomorphic Vector Spaces*

Invertible Linear Maps

We begin this section by defining the notions of invertible and inverse in the context of linear maps.

3.53 Definition ***invertible, inverse***

- A linear map $T \in \mathcal{L}(V, W)$ is called ***invertible*** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that ST equals the identity map on V and TS equals the identity map on W.
- A linear map $S \in \mathcal{L}(W, V)$ satisfying $ST = I$ and $TS = I$ is called an ***inverse*** of T (note that the first I is the identity map on V and the second I is the identity map on W).

3.54 Inverse is unique

An invertible linear map has a unique inverse.

Proof Suppose $T \in \mathcal{L}(V, W)$ is invertible and S_1 and S_2 are inverses of T. Then

$$S_1 = S_1 I = S_1(TS_2) = (S_1 T)S_2 = IS_2 = S_2.$$

Thus $S_1 = S_2$. ■

Now that we know that the inverse is unique, we can give it a notation.

3.55 Notation T^{-1}

If T is invertible, then its inverse is denoted by T^{-1}. In other words, if $T \in \mathcal{L}(V, W)$ is invertible, then T^{-1} is the unique element of $\mathcal{L}(W, V)$ such that $T^{-1}T = I$ and $TT^{-1} = I$.

The following result characterizes the invertible linear maps.

3.56 Invertibility is equivalent to injectivity and surjectivity

A linear map is invertible if and only if it is injective and surjective.

Proof Suppose $T \in \mathcal{L}(V, W)$. We need to show that T is invertible if and only if it is injective and surjective.

First suppose T is invertible. To show that T is injective, suppose $u, v \in V$ and $Tu = Tv$. Then

$$u = T^{-1}(Tu) = T^{-1}(Tv) = v,$$

so $u = v$. Hence T is injective.

We are still assuming that T is invertible. Now we want to prove that T is surjective. To do this, let $w \in W$. Then $w = T(T^{-1}w)$, which shows that w is in the range of T. Thus range $T = W$. Hence T is surjective, completing this direction of the proof.

Now suppose T is injective and surjective. We want to prove that T is invertible. For each $w \in W$, define Sw to be the unique element of V such that $T(Sw) = w$ (the existence and uniqueness of such an element follow from the surjectivity and injectivity of T). Clearly $T \circ S$ equals the identity map on W.

To prove that $S \circ T$ equals the identity map on V, let $v \in V$. Then

$$T\big((S \circ T)v\big) = (T \circ S)(Tv) = I(Tv) = Tv.$$

This equation implies that $(S \circ T)v = v$ (because T is injective). Thus $S \circ T$ equals the identity map on V.

To complete the proof, we need to show that S is linear. To do this, suppose $w_1, w_2 \in W$. Then

$$T(Sw_1 + Sw_2) = T(Sw_1) + T(Sw_2) = w_1 + w_2.$$

Thus $Sw_1 + Sw_2$ is the unique element of V that T maps to $w_1 + w_2$. By the definition of S, this implies that $S(w_1 + w_2) = Sw_1 + Sw_2$. Hence S satisfies the additive property required for linearity.

The proof of homogeneity is similar. Specifically, if $w \in W$ and $\lambda \in \mathbf{F}$, then

$$T(\lambda Sw) = \lambda T(Sw) = \lambda w.$$

Thus λSw is the unique element of V that T maps to λw. By the definition of S, this implies that $S(\lambda w) = \lambda Sw$. Hence S is linear, as desired. ■

3.57 Example ***linear maps that are not invertible***

- The multiplication by x^2 linear map from $\mathcal{P}(\mathbf{R})$ to $\mathcal{P}(\mathbf{R})$ (see 3.4) is not invertible because it is not surjective (1 is not in the range).
- The backward shift linear map from $\mathbf{F}^\infty$ to $\mathbf{F}^\infty$ (see 3.4) is not invertible because it is not injective [$(1, 0, 0, 0, \ldots)$ is in the null space].

Isomorphic Vector Spaces

The next definition captures the idea of two vector spaces that are essentially the same, except for the names of the elements of the vector spaces.

3.58 Definition ***isomorphism, isomorphic***

- An ***isomorphism*** is an invertible linear map.
- Two vector spaces are called ***isomorphic*** if there is an isomorphism from one vector space onto the other one.

Think of an isomorphism $T: V \to W$ as relabeling $v \in V$ as $Tv \in W$. This viewpoint explains why two isomorphic vector spaces have the same vector space properties. The terms "isomorphism" and "invertible linear map" mean the same thing. Use "isomorphism" when you want to emphasize that the two spaces are essentially the same.

*The Greek word **isos** means equal; the Greek word **morph** means shape. Thus **isomorphic** literally means equal shape.*

3.59 Dimension shows whether vector spaces are isomorphic

Two finite-dimensional vector spaces over **F** are isomorphic if and only if they have the same dimension.

Proof First suppose V and W are isomorphic finite-dimensional vector spaces. Thus there exists an isomorphism T from V onto W. Because T is invertible, we have $\operatorname{null} T = \{0\}$ and $\operatorname{range} T = W$. Thus $\dim \operatorname{null} T = 0$ and $\dim \operatorname{range} T = \dim W$. The formula

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$$

(the Fundamental Theorem of Linear Maps, which is 3.22) thus becomes the equation $\dim V = \dim W$, completing the proof in one direction.

To prove the other direction, suppose V and W are finite-dimensional vector spaces with the same dimension. Let $v_1, \dots, v_n$ be a basis of V and $w_1, \dots, w_n$ be a basis of W. Let $T \in \mathcal{L}(V, W)$ be defined by

$$T(c_1 v_1 + \cdots + c_n v_n) = c_1 w_1 + \cdots + c_n w_n.$$

Then T is a well-defined linear map because $v_1, \dots, v_n$ is a basis of V (see 3.5). Also, T is surjective because $w_1, \dots, w_n$ spans W. Furthermore, $\operatorname{null} T = \{0\}$ because $w_1, \dots, w_n$ is linearly independent; thus T is injective. Because T is injective and surjective, it is an isomorphism (see 3.56). Hence V and W are isomorphic, as desired. ■

The previous result implies that each finite-dimensional vector space V is isomorphic to $\mathbf{F}^n$, where $n = \dim V$.

Because every finite-dimensional vector space is isomorphic to some $\mathbf{F}^n$, why not just study $\mathbf{F}^n$ instead of more general vector spaces? To answer this question, note that an investigation of $\mathbf{F}^n$ would soon lead to other vector spaces. For example, we would encounter the null space and range of linear maps. Although each of these vector spaces is isomorphic to some $\mathbf{F}^n$, thinking of them that way often adds complexity but no new insight.

If $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W, then for each $T \in \mathcal{L}(V, W)$, we have a matrix $\mathcal{M}(T) \in \mathbf{F}^{m,n}$. In other words, once bases have been fixed for V and W, $\mathcal{M}$ becomes a function from $\mathcal{L}(V, W)$ to $\mathbf{F}^{m,n}$. Notice that 3.36 and 3.38 show that $\mathcal{M}$ is a linear map. This linear map is actually invertible, as we now show.

3.60 $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$ are isomorphic

Suppose $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. Then $\mathcal{M}$ is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$.

Proof We already noted that $\mathcal{M}$ is linear. We need to prove that $\mathcal{M}$ is injective and surjective. Both are easy. We begin with injectivity. If $T \in \mathcal{L}(V, W)$ and $\mathcal{M}(T) = 0$, then $Tv_k = 0$ for $k = 1, \dots, n$. Because $v_1, \dots, v_n$ is a basis of V, this implies $T = 0$. Thus $\mathcal{M}$ is injective (by 3.16).

To prove that $\mathcal{M}$ is surjective, suppose $A \in \mathbf{F}^{m,n}$. Let T be the linear map from V to W such that

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j$$

for $k = 1, \dots, n$ (see 3.5). Obviously $\mathcal{M}(T)$ equals A, and thus the range of $\mathcal{M}$ equals $\mathbf{F}^{m,n}$, as desired. ■

Now we can determine the dimension of the vector space of linear maps from one finite-dimensional vector space to another.

3.61 $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$

Suppose V and W are finite-dimensional. Then $\mathcal{L}(V, W)$ is finite-dimensional and

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Proof This follows from 3.60, 3.59, and 3.40. ■

Linear Maps Thought of as Matrix Multiplication

Previously we defined the matrix of a linear map. Now we define the matrix of a vector.

3.62 Definition ***matrix of a vector, $\mathcal{M}(v)$***

Suppose $v \in V$ and $v_1, \dots, v_n$ is a basis of V. The ***matrix of*** v with respect to this basis is the n-by-1 matrix

$$\mathcal{M}(v) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

where $c_1, \dots, c_n$ are the scalars such that

$$v = c_1 v_1 + \cdots + c_n v_n.$$

The matrix $\mathcal{M}(v)$ of a vector $v \in V$ depends on the basis $v_1, \dots, v_n$ of V, as well as on v. However, the basis should be clear from the context and thus it is not included in the notation.

3.63 Example ***matrix of a vector***

- The matrix of $2 - 7x + 5x^3$ with respect to the standard basis of $\mathcal{P}_3(\mathbf{R})$ is

$$\begin{pmatrix} 2 \\ -7 \\ 0 \\ 5 \end{pmatrix}.$$

- The matrix of a vector $x \in \mathbf{F}^n$ with respect to the standard basis is obtained by writing the coordinates of x as the entries in an n-by-1 matrix. In other words, if $x = (x_1, \dots, x_n) \in \mathbf{F}^n$, then

$$\mathcal{M}(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Occasionally we want to think of elements of V as relabeled to be n-by-1 matrices. Once a basis $v_1, \dots, v_n$ is chosen, the function $\mathcal{M}$ that takes $v \in V$ to $\mathcal{M}(v)$ is an isomorphism of V onto $\mathbf{F}^{n,1}$ that implements this relabeling.

Recall that if A is an m-by-n matrix, then $A_{\cdot,k}$ denotes the k^{th} column of A, thought of as an m-by-1 matrix. In the next result, $\mathcal{M}(v_k)$ is computed with respect to the basis $w_1, \ldots, w_m$ of W.

3.64 $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(v_k)$.

Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Let $1 \leq k \leq n$. Then the k^{th} column of $\mathcal{M}(T)$, which is denoted by $\mathcal{M}(T)_{\cdot,k}$, equals $\mathcal{M}(v_k)$.

Proof The desired result follows immediately from the definitions of $\mathcal{M}(T)$ and $\mathcal{M}(v_k)$. ■

The next result shows how the notions of the matrix of a linear map, the matrix of a vector, and matrix multiplication fit together.

3.65 Linear maps act like matrix multiplication

Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

Proof Suppose $v = c_1 v_1 + \cdots + c_n v_n$, where $c_1, \ldots, c_n \in \mathbf{F}$. Thus

3.66 $$Tv = c_1 T v_1 + \cdots + c_n T v_n.$$

Hence

$$\begin{aligned} \mathcal{M}(Tv) &= c_1 \mathcal{M}(Tv_1) + \cdots + c_n \mathcal{M}(Tv_n) \\ &= c_1 \mathcal{M}(T)_{\cdot,1} + \cdots + c_n \mathcal{M}(T)_{\cdot,n} \\ &= \mathcal{M}(T)\mathcal{M}(v), \end{aligned}$$

where the first equality follows from 3.66 and the linearity of $\mathcal{M}$, the second equality comes from 3.64, and the last equality comes from 3.52. ■

Each m-by-n matrix A induces a linear map from $\mathbf{F}^{n,1}$ to $\mathbf{F}^{m,1}$, namely the matrix multiplication function that takes $x \in \mathbf{F}^{n,1}$ to $Ax \in \mathbf{F}^{m,1}$. The result above can be used to think of every linear map (from one finite-dimensional vector space to another finite-dimensional vector space) as a matrix multiplication map after suitable relabeling via the isomorphisms given by $\mathcal{M}$. Specifically, if $T \in \mathcal{L}(V, W)$ and we identify $v \in V$ with $\mathcal{M}(v) \in \mathbf{F}^{n,1}$, then the result above says that we can identify Tv with $\mathcal{M}(T)\mathcal{M}(v)$.

Because the result above allows us to think (via isomorphisms) of each linear map as multiplication on $\mathbf{F}^{n,1}$ by some matrix A, keep in mind that the specific matrix A depends not only on the linear map but also on the choice of bases. One of the themes of many of the most important results in later chapters will be the choice of a basis that makes the matrix A as simple as possible.

In this book, we concentrate on linear maps rather than on matrices. However, sometimes thinking of linear maps as matrices (or thinking of matrices as linear maps) gives important insights that we will find useful.

Operators

Linear maps from a vector space to itself are so important that they get a special name and special notation.

3.67 Definition ***operator,*** $\mathcal{L}(V)$

- A linear map from a vector space to itself is called an ***operator***.
- The notation $\mathcal{L}(V)$ denotes the set of all operators on V. In other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

The deepest and most important parts of linear algebra, as well as most of the rest of this book, deal with operators.

A linear map is invertible if it is injective and surjective. For an operator, you might wonder whether injectivity alone, or surjectivity alone, is enough to imply invertibility. On infinite-dimensional vector spaces, neither condition alone implies invertibility, as illustrated by the next example, which uses two familiar operators from Example 3.4.

3.68 Example ***neither injectivity nor surjectivity implies invertibility***

- The multiplication by x^2 operator on $\mathcal{P}(\mathbf{R})$ is injective but not surjective.
- The backward shift operator on $\mathbf{F}^\infty$ is surjective but not injective.

In view of the example above, the next result is remarkable—it states that for operators on a finite-dimensional vector space, either injectivity or surjectivity alone implies the other condition. Often it is easier to check that an operator on a finite-dimensional vector space is injective, and then we get surjectivity for free.

3.69 Injectivity is equivalent to surjectivity in finite dimensions

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is invertible;

(b) T is injective;

(c) T is surjective.

Proof Clearly (a) implies (b).

Now suppose (b) holds, so that T is injective. Thus null $T = \{0\}$ (by 3.16). From the Fundamental Theorem of Linear Maps (3.22) we have

$$\begin{aligned}\dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T \\ &= \dim V.\end{aligned}$$

Thus range T equals V. Thus T is surjective. Hence (b) implies (c).

Now suppose (c) holds, so that T is surjective. Thus range $T = V$. From the Fundamental Theorem of Linear Maps (3.22) we have

$$\begin{aligned}\dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &= 0.\end{aligned}$$

Thus null T equals $\{0\}$. Thus T is injective (by 3.16), and so T is invertible (we already knew that T was surjective). Hence (c) implies (a), completing the proof. ■

The next example illustrates the power of the previous result. Although it is possible to prove the result in the example below without using linear algebra, the proof using linear algebra is cleaner and easier.

3.70 **Example** Show that for each polynomial $q \in \mathcal{P}(\mathbf{R})$, there exists a polynomial $p \in \mathcal{P}(\mathbf{R})$ with $\big((x^2 + 5x + 7)p\big)'' = q$.

Solution Example 3.68 shows that the magic of 3.69 does not apply to the infinite-dimensional vector space $\mathcal{P}(\mathbf{R})$. However, each nonzero polynomial q has some degree m. By restricting attention to $\mathcal{P}_m(\mathbf{R})$, we can work with a finite-dimensional vector space.

Suppose $q \in \mathcal{P}_m(\mathbf{R})$. Define $T : \mathcal{P}_m(\mathbf{R}) \to \mathcal{P}_m(\mathbf{R})$ by

$$Tp = \big((x^2 + 5x + 7)p\big)''.$$

Multiplying a nonzero polynomial by $(x^2 + 5x + 7)$ increases the degree by 2, and then differentiating twice reduces the degree by 2. Thus T is indeed an operator on $\mathcal{P}_m(\mathbf{R})$.

Every polynomial whose second derivative equals 0 is of the form $ax + b$, where $a, b \in \mathbf{R}$. Thus $\operatorname{null} T = \{0\}$. Hence T is injective.

Now 3.69 implies that T is surjective. Thus there exists a polynomial $p \in \mathcal{P}_m(\mathbf{R})$ such that $\big((x^2 + 5x + 7)p\big)'' = q$, as desired.

Exercise 30 in Section 6.A gives a similar but more spectacular application of 3.69. The result in that exercise is quite difficult to prove without using linear algebra.

EXERCISES 3.D

1 Suppose $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$ are both invertible linear maps. Prove that $ST \in \mathcal{L}(U, W)$ is invertible and that $(ST)^{-1} = T^{-1}S^{-1}$.

2 Suppose V is finite-dimensional and $\dim V > 1$. Prove that the set of noninvertible operators on V is not a subspace of $\mathcal{L}(V)$.

3 Suppose V is finite-dimensional, U is a subspace of V, and $S \in \mathcal{L}(U, V)$. Prove there exists an invertible operator $T \in \mathcal{L}(V)$ such that $Tu = Su$ for every $u \in U$ if and only if S is injective.

4 Suppose W is finite-dimensional and $T_1, T_2 \in \mathcal{L}(V, W)$. Prove that $\operatorname{null} T_1 = \operatorname{null} T_2$ if and only if there exists an invertible operator $S \in \mathcal{L}(W)$ such that $T_1 = ST_2$.

5 Suppose V is finite-dimensional and $T_1, T_2 \in \mathcal{L}(V, W)$. Prove that $\operatorname{range} T_1 = \operatorname{range} T_2$ if and only if there exists an invertible operator $S \in \mathcal{L}(V)$ such that $T_1 = T_2 S$.

6 Suppose V and W are finite-dimensional and $T_1, T_2 \in \mathcal{L}(V, W)$. Prove that there exist invertible operators $R \in \mathcal{L}(V)$ and $S \in \mathcal{L}(W)$ such that $T_1 = ST_2R$ if and only if $\dim \operatorname{null} T_1 = \dim \operatorname{null} T_2$.

7 Suppose V and W are finite-dimensional. Let $v \in V$. Let

$$E = \{T \in \mathcal{L}(V, W) : Tv = 0\}.$$

(a) Show that E is a subspace of $\mathcal{L}(V, W)$.

(b) Suppose $v \neq 0$. What is $\dim E$?

8 Suppose V is finite-dimensional and $T: V \to W$ is a surjective linear map of V onto W. Prove that there is a subspace U of V such that $T|_U$ is an isomorphism of U onto W. (Here $T|_U$ means the function T restricted to U. In other words, $T|_U$ is the function whose domain is U, with $T|_U$ defined by $T|_U(u) = Tu$ for every $u \in U$.)

9 Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST is invertible if and only if both S and T are invertible.

10 Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that $ST = I$ if and only if $TS = I$.

11 Suppose V is finite-dimensional and $S, T, U \in \mathcal{L}(V)$ and $STU = I$. Show that T is invertible and that $T^{-1} = US$.

12 Show that the result in the previous exercise can fail without the hypothesis that V is finite-dimensional.

13 Suppose V is a finite-dimensional vector space and $R, S, T \in \mathcal{L}(V)$ are such that RST is surjective. Prove that S is injective.

14 Suppose $v_1, \dots, v_n$ is a basis of V. Prove that the map $T: V \to \mathbf{F}^{n,1}$ defined by

$$Tv = \mathcal{M}(v)$$

is an isomorphism of V onto $\mathbf{F}^{n,1}$; here $\mathcal{M}(v)$ is the matrix of $v \in V$ with respect to the basis $v_1, \dots, v_n$.

15 Prove that every linear map from $\mathbf{F}^{n,1}$ to $\mathbf{F}^{m,1}$ is given by a matrix multiplication. In other words, prove that if $T \in \mathcal{L}(\mathbf{F}^{n,1}, \mathbf{F}^{m,1})$, then there exists an m-by-n matrix A such that $Tx = Ax$ for every $x \in \mathbf{F}^{n,1}$.

16 Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that T is a scalar multiple of the identity if and only if $ST = TS$ for every $S \in \mathcal{L}(V)$.

17 Suppose V is finite-dimensional and $\mathcal{E}$ is a subspace of $\mathcal{L}(V)$ such that $ST \in \mathcal{E}$ and $TS \in \mathcal{E}$ for all $S \in \mathcal{L}(V)$ and all $T \in \mathcal{E}$. Prove that $\mathcal{E} = \{0\}$ or $\mathcal{E} = \mathcal{L}(V)$.

18 Show that V and $\mathcal{L}(\mathbf{F}, V)$ are isomorphic vector spaces.

19 Suppose $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R})\big)$ is such that T is injective and $\deg Tp \le \deg p$ for every nonzero polynomial $p \in \mathcal{P}(\mathbf{R})$.

(a) Prove that T is surjective.

(b) Prove that $\deg Tp = \deg p$ for every nonzero $p \in \mathcal{P}(\mathbf{R})$.

20 Suppose n is a positive integer and $A_{i,j} \in \mathbf{F}$ for $i, j = 1, \dots, n$. Prove that the following are equivalent (note that in both parts below, the number of equations equals the number of variables):

(a) The trivial solution $x_1 = \cdots = x_n = 0$ is the only solution to the homogeneous system of equations

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= 0 \\ &\vdots \\ \sum_{k=1}^{n} A_{n,k} x_k &= 0. \end{aligned}$$

(b) For every $c_1, \dots, c_n \in \mathbf{F}$, there exists a solution to the system of equations

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= c_1 \\ &\vdots \\ \sum_{k=1}^{n} A_{n,k} x_k &= c_n. \end{aligned}$$

3.E Products and Quotients of Vector Spaces

Products of Vector Spaces

As usual when dealing with more than one vector space, all the vector spaces in use should be over the same field.

3.71 Definition ***product of vector spaces***

Suppose $V_1, \ldots, V_m$ are vector spaces over $\mathbf{F}$.

- The ***product*** $V_1 \times \cdots \times V_m$ is defined by
$$V_1 \times \cdots \times V_m = \{(v_1, \ldots, v_m) : v_1 \in V_1, \ldots, v_m \in V_m\}.$$
- Addition on $V_1 \times \cdots \times V_m$ is defined by
$$(u_1, \ldots, u_m) + (v_1, \ldots, v_m) = (u_1 + v_1, \ldots, u_m + v_m).$$
- Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by
$$\lambda(v_1, \ldots, v_m) = (\lambda v_1, \ldots, \lambda v_m).$$

3.72 Example Elements of $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^3$ are lists of length 2, with the first item in the list an element of $\mathcal{P}_2(\mathbf{R})$ and the second item in the list an element of $\mathbf{R}^3$.

For example, $\big(5 - 6x + 4x^2, (3, 8, 7)\big) \in \mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^3$.

The next result should be interpreted to mean that the product of vector spaces is a vector space with the operations of addition and scalar multiplication as defined above.

3.73 Product of vector spaces is a vector space

Suppose $V_1, \ldots, V_m$ are vector spaces over $\mathbf{F}$. Then $V_1 \times \cdots \times V_m$ is a vector space over $\mathbf{F}$.

The proof of the result above is left to the reader. Note that the additive identity of $V_1 \times \cdots \times V_m$ is $(0, \ldots, 0)$, where the 0 in the j^{th} slot is the additive identity of V_j. The additive inverse of $(v_1, \ldots, v_m) \in V_1 \times \cdots \times V_m$ is $(-v_1, \ldots, -v_m)$.

3.74 Example Is $\mathbf{R}^2 \times \mathbf{R}^3$ equal to $\mathbf{R}^5$? Is $\mathbf{R}^2 \times \mathbf{R}^3$ isomorphic to $\mathbf{R}^5$?

Solution Elements of $\mathbf{R}^2 \times \mathbf{R}^3$ are lists $\big((x_1, x_2), (x_3, x_4, x_5)\big)$, where $x_1, x_2, x_3, x_4, x_5 \in \mathbf{R}$.

Elements of $\mathbf{R}^5$ are lists $(x_1, x_2, x_3, x_4, x_5)$, where $x_1, x_2, x_3, x_4, x_5 \in \mathbf{R}$.

Although these look almost the same, they are not the same kind of object. Elements of $\mathbf{R}^2 \times \mathbf{R}^3$ are lists of length 2 (with the first item itself a list of length 2 and the second item a list of length 3), and elements of $\mathbf{R}^5$ are lists of length 5. Thus $\mathbf{R}^2 \times \mathbf{R}^3$ does not equal $\mathbf{R}^5$.

The linear map that takes a vector $\big((x_1, x_2), (x_3, x_4, x_5)\big) \in \mathbf{R}^2 \times \mathbf{R}^3$ to $(x_1, x_2, x_3, x_4, x_5) \in \mathbf{R}^5$ is clearly an isomorphism of $\mathbf{R}^2 \times \mathbf{R}^3$ onto $\mathbf{R}^5$. Thus these two vector spaces are isomorphic.

In this case, the isomorphism is so natural that we should think of it as a relabeling. Some people would even informally say that $\mathbf{R}^2 \times \mathbf{R}^3$ equals $\mathbf{R}^5$, which is not technically correct but which captures the spirit of identification via relabeling.

The next example illustrates the idea of the proof of 3.76.

3.75 Example Find a basis of $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$.

Solution Consider this list of length 5 of elements of $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$:

$$\big(1, (0, 0)\big), \big(x, (0, 0)\big), \big(x^2, (0, 0)\big), \big(0, (1, 0)\big), \big(0, (0, 1)\big).$$

The list above is linearly independent and it spans $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$. Thus it is a basis of $\mathcal{P}_2(\mathbf{R}) \times \mathbf{R}^2$.

3.76 Dimension of a product is the sum of dimensions

Suppose $V_1, \dots, V_m$ are finite-dimensional vector spaces. Then $V_1 \times \dots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \dots \times V_m) = \dim V_1 + \dots + \dim V_m.$$

Proof Choose a basis of each V_j. For each basis vector of each V_j, consider the element of $V_1 \times \dots \times V_m$ that equals the basis vector in the j^{th} slot and 0 in the other slots. The list of all such vectors is linearly independent and spans $V_1 \times \dots \times V_m$. Thus it is a basis of $V_1 \times \dots \times V_m$. The length of this basis is $\dim V_1 + \dots + \dim V_m$, as desired. ∎

Products and Direct Sums

In the next result, the map Γ is surjective by the definition of $U_1 + \cdots + U_m$. Thus the last word in the result below could be changed from "injective" to "invertible".

3.77 Products and direct sums

Suppose that $U_1, \ldots, U_m$ are subspaces of V. Define a linear map $\Gamma : U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by

$$\Gamma(u_1, \ldots, u_m) = u_1 + \cdots + u_m.$$

Then $U_1 + \cdots + U_m$ is a direct sum if and only if Γ is injective.

Proof The linear map Γ is injective if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0. Thus 1.44 shows that Γ is injective if and only if $U_1 + \cdots + U_m$ is a direct sum, as desired. ■

3.78 A sum is a direct sum if and only if dimensions add up

Suppose V is finite-dimensional and $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m.$$

Proof The map Γ in 3.77 is surjective. Thus by the Fundamental Theorem of Linear Maps (3.22), Γ is injective if and only if

$$\dim(U_1 + \cdots + U_m) = \dim(U_1 \times \cdots \times U_m).$$

Combining 3.77 and 3.76 now shows that $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m,$$

as desired. ■

In the special case $m = 2$, an alternative proof that $U_1 + U_2$ is a direct sum if and only if $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$ can be obtained by combining 1.45 and 2.43.

Quotients of Vector Spaces

We begin our approach to quotient spaces by defining the sum of a vector and a subspace.

3.79 Definition $v + U$

Suppose $v \in V$ and U is a subspace of V. Then $v + U$ is the subset of V defined by

$$v + U = \{v + u : u \in U\}.$$

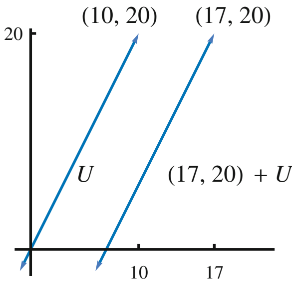
3.80 Example Suppose

$$U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}.$$

Then U is the line in $\mathbf{R}^2$ through the origin with slope 2. Thus

$$(17, 20) + U$$

is the line in $\mathbf{R}^2$ that contains the point $(17, 20)$ and has slope 2.



3.81 Definition ***affine subset, parallel***

- An ***affine subset*** of V is a subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V.
- For $v \in V$ and U a subspace of V, the affine subset $v + U$ is said to be ***parallel*** to U.

3.82 Example ***parallel affine subsets***

- In Example 3.80 above, all the lines in $\mathbf{R}^2$ with slope 2 are parallel to U.
- If $U = \{(x, y, 0) \in \mathbf{R}^3 : x, y \in \mathbf{R}\}$, then the affine subsets of $\mathbf{R}^3$ parallel to U are the planes in $\mathbf{R}^3$ that are parallel to the xy-plane U in the usual sense.
  Important: With the definition of ***parallel*** given in 3.81, no line in $\mathbf{R}^3$ is considered to be an affine subset that is parallel to the plane U.

3.83 Definition ***quotient space,*** V/U

Suppose U is a subspace of V. Then the ***quotient space*** V/U is the set of all affine subsets of V parallel to U. In other words,

$$V/U = \{v + U : v \in V\}.$$

3.84 Example ***quotient spaces***

- If $U = \{(x, 2x) \in \mathbf{R}^2 : x \in \mathbf{R}\}$, then $\mathbf{R}^2/U$ is the set of all lines in $\mathbf{R}^2$ that have slope 2.
- If U is a line in $\mathbf{R}^3$ containing the origin, then $\mathbf{R}^3/U$ is the set of all lines in $\mathbf{R}^3$ parallel to U.
- If U is a plane in $\mathbf{R}^3$ containing the origin, then $\mathbf{R}^3/U$ is the set of all planes in $\mathbf{R}^3$ parallel to U.

Our next goal is to make V/U into a vector space. To do this, we will need the following result.

3.85 Two affine subsets parallel to U are equal or disjoint

Suppose U is a subspace of V and $v, w \in V$. Then the following are equivalent:

(a) $v - w \in U$;

(b) $v + U = w + U$;

(c) $(v + U) \cap (w + U) \neq \varnothing$.

Proof First suppose (a) holds, so $v - w \in U$. If $u \in U$, then

$$v + u = w + \big((v - w) + u\big) \in w + U.$$

Thus $v + U \subset w + U$. Similarly, $w + U \subset v + U$. Thus $v + U = w + U$, completing the proof that (a) implies (b).

Obviously (b) implies (c).

Now suppose (c) holds, so $(v + U) \cap (w + U) \neq \varnothing$. Thus there exist $u_1, u_2 \in U$ such that

$$v + u_1 = w + u_2.$$

Thus $v - w = u_2 - u_1$. Hence $v - w \in U$, showing that (c) implies (a) and completing the proof. ■

Now we can define addition and scalar multiplication on V/U.

3.86 Definition ***addition and scalar multiplication on*** V/U

Suppose U is a subspace of V. Then ***addition*** and ***scalar multiplication*** are defined on V/U by

$$(v+U)+(w+U)=(v+w)+U$$
$$\lambda(v+U)=(\lambda v)+U$$

for $v, w \in V$ and $\lambda \in \mathbf{F}$.

As part of the proof of the next result, we will show that the definitions above make sense.

3.87 Quotient space is a vector space

Suppose U is a subspace of V. Then V/U, with the operations of addition and scalar multiplication as defined above, is a vector space.

Proof The potential problem with the definitions above of addition and scalar multiplication on V/U is that the representation of an affine subset parallel to U is not unique. Specifically, suppose $v, w \in V$. Suppose also that $\hat{v}, \hat{w} \in V$ are such that $v+U=\hat{v}+U$ and $w+U=\hat{w}+U$. To show that the definition of addition on V/U given above makes sense, we must show that $(v+w)+U=(\hat{v}+\hat{w})+U$.

By 3.85, we have

$$v-\hat{v} \in U \quad \text{and} \quad w-\hat{w} \in U.$$

Because U is a subspace of V and thus is closed under addition, this implies that $(v-\hat{v})+(w-\hat{w}) \in U$. Thus $(v+w)-(\hat{v}+\hat{w}) \in U$. Using 3.85 again, we see that

$$(v+w)+U=(\hat{v}+\hat{w})+U,$$

as desired. Thus the definition of addition on V/U makes sense.

Similarly, suppose $\lambda \in \mathbf{F}$. Because U is a subspace of V and thus is closed under scalar multiplication, we have $\lambda(v-\hat{v}) \in U$. Thus $\lambda v-\lambda\hat{v} \in U$. Hence 3.85 implies that $(\lambda v)+U=(\lambda\hat{v})+U$. Thus the definition of scalar multiplication on V/U makes sense.

Now that addition and scalar multiplication have been defined on V/U, the verification that these operations make V/U into a vector space is straightforward and is left to the reader. Note that the additive identity of V/U is $0+U$ (which equals U) and that the additive inverse of $v+U$ is $(-v)+U$. ■

The next concept will give us an easy way to compute the dimension of V/U.

3.88 Definition ***quotient map, π***

Suppose U is a subspace of V. The ***quotient map*** π is the linear map $\pi: V \to V/U$ defined by

$$\pi(v) = v + U$$

for $v \in V$.

The reader should verify that π is indeed a linear map. Although π depends on U as well as V, these spaces are left out of the notation because they should be clear from the context.

3.89 Dimension of a quotient space

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim V/U = \dim V - \dim U.$$

Proof Let π be the quotient map from V to V/U. From 3.85, we see that $\operatorname{null} \pi = U$. Clearly $\operatorname{range} \pi = V/U$. The Fundamental Theorem of Linear Maps (3.22) thus tells us that

$$\dim V = \dim U + \dim V/U,$$

which gives the desired result. ■

Each linear map T on V induces a linear map $\tilde{T}$ on $V/(\operatorname{null} T)$, which we now define.

3.90 Definition $\tilde{T}$

Suppose $T \in \mathcal{L}(V, W)$. Define $\tilde{T}: V/(\operatorname{null} T) \to W$ by

$$\tilde{T}(v + \operatorname{null} T) = Tv.$$

To show that the definition of $\tilde{T}$ makes sense, suppose $u, v \in V$ are such that $u + \operatorname{null} T = v + \operatorname{null} T$. By 3.85, we have $u - v \in \operatorname{null} T$. Thus $T(u - v) = 0$. Hence $Tu = Tv$. Thus the definition of $\tilde{T}$ indeed makes sense.

3.91 Null space and range of $\tilde{T}$

Suppose $T \in \mathcal{L}(V, W)$. Then

(a) $\tilde{T}$ is a linear map from $V/(\operatorname{null} T)$ to W;

(b) $\tilde{T}$ is injective;

(c) $\operatorname{range} \tilde{T} = \operatorname{range} T$;

(d) $V/(\operatorname{null} T)$ is isomorphic to $\operatorname{range} T$.

Proof

(a) The routine verification that $\tilde{T}$ is linear is left to the reader.

(b) Suppose $v \in V$ and $\tilde{T}(v + \operatorname{null} T) = 0$. Then $Tv = 0$. Thus $v \in \operatorname{null} T$. Hence 3.85 implies that $v + \operatorname{null} T = 0 + \operatorname{null} T$. This implies that $\operatorname{null} \tilde{T} = 0$, and hence $\tilde{T}$ is injective, as desired.

(c) The definition of $\tilde{T}$ shows that $\operatorname{range} \tilde{T} = \operatorname{range} T$.

(d) Parts (b) and (c) imply that if we think of $\tilde{T}$ as mapping into $\operatorname{range} T$, then $\tilde{T}$ is an isomorphism from $V/(\operatorname{null} T)$ onto $\operatorname{range} T$. ■

EXERCISES 3.E

1 Suppose T is a function from V to W. The ***graph*** of T is the subset of $V \times W$ defined by

$$\text{graph of } T = \{(v, Tv) \in V \times W : v \in V\}.$$

Prove that T is a linear map if and only if the graph of T is a subspace of $V \times W$.

[*Formally, a function T from V to W is a subset T of $V \times W$ such that for each $v \in V$, there exists exactly one element $(v, w) \in T$. In other words, formally a function is what is called above its graph. We do not usually think of functions in this formal manner. However, if we do become formal, then the exercise above could be rephrased as follows: Prove that a function T from V to W is a linear map if and only if T is a subspace of $V \times W$.*]

2 Suppose $V_1, \dots, V_m$ are vector spaces such that $V_1 \times \cdots \times V_m$ is finite-dimensional. Prove that V_j is finite-dimensional for each $j = 1, \dots, m$.

3 Give an example of a vector space V and subspaces U_1, U_2 of V such that $U_1 \times U_2$ is isomorphic to $U_1 + U_2$ but $U_1 + U_2$ is not a direct sum.

Hint: The vector space V must be infinite-dimensional.

4 Suppose $V_1, \dots, V_m$ are vector spaces. Prove that $\mathcal{L}(V_1 \times \cdots \times V_m, W)$ and $\mathcal{L}(V_1, W) \times \cdots \times \mathcal{L}(V_m, W)$ are isomorphic vector spaces.

5 Suppose $W_1, \dots, W_m$ are vector spaces. Prove that $\mathcal{L}(V, W_1 \times \cdots \times W_m)$ and $\mathcal{L}(V, W_1) \times \cdots \times \mathcal{L}(V, W_m)$ are isomorphic vector spaces.

6 For n a positive integer, define V^n by

$$V^n = \underbrace{V \times \cdots \times V}_{n \text{ times}}.$$

Prove that V^n and $\mathcal{L}(\mathbf{F}^n, V)$ are isomorphic vector spaces.

7 Suppose v, x are vectors in V and U, W are subspaces of V such that $v + U = x + W$. Prove that $U = W$.

8 Prove that a nonempty subset A of V is an affine subset of V if and only if $\lambda v + (1 - \lambda)w \in A$ for all $v, w \in A$ and all $\lambda \in \mathbf{F}$.

9 Suppose A_1 and A_2 are affine subsets of V. Prove that the intersection $A_1 \cap A_2$ is either an affine subset of V or the empty set.

10 Prove that the intersection of every collection of affine subsets of V is either an affine subset of V or the empty set.

11 Suppose $v_1, \dots, v_m \in V$. Let

$$A = \{\lambda_1 v_1 + \cdots + \lambda_m v_m : \lambda_1, \dots, \lambda_m \in \mathbf{F} \text{ and } \lambda_1 + \cdots + \lambda_m = 1\}.$$

(a) Prove that A is an affine subset of V.

(b) Prove that every affine subset of V that contains $v_1, \dots, v_m$ also contains A.

(c) Prove that $A = v + U$ for some $v \in V$ and some subspace U of V with $\dim U \leq m - 1$.

12 Suppose U is a subspace of V such that V/U is finite-dimensional. Prove that V is isomorphic to $U \times (V/U)$.

13 Suppose U is a subspace of V and $v_1 + U, \ldots, v_m + U$ is a basis of V/U and $u_1, \ldots, u_n$ is a basis of U. Prove that $v_1, \ldots, v_m, u_1, \ldots, u_n$ is a basis of V.

14 Suppose $U = \{(x_1, x_2, \ldots) \in \mathbf{F}^\infty : x_j \neq 0 \text{ for only finitely many } j\}$.

(a) Show that U is a subspace of $\mathbf{F}^\infty$.

(b) Prove that $\mathbf{F}^\infty/U$ is infinite-dimensional.

15 Suppose $\varphi \in \mathcal{L}(V, \mathbf{F})$ and $\varphi \neq 0$. Prove that $\dim V/(\operatorname{null} \varphi) = 1$.

16 Suppose U is a subspace of V such that $\dim V/U = 1$. Prove that there exists $\varphi \in \mathcal{L}(V, \mathbf{F})$ such that $\operatorname{null} \varphi = U$.

17 Suppose U is a subspace of V such that V/U is finite-dimensional. Prove that there exists a subspace W of V such that $\dim W = \dim V/U$ and $V = U \oplus W$.

18 Suppose $T \in \mathcal{L}(V, W)$ and U is a subspace of V. Let π denote the quotient map from V onto V/U. Prove that there exists $S \in \mathcal{L}(V/U, W)$ such that $T = S \circ \pi$ if and only if $U \subset \operatorname{null} T$.

19 Find a correct statement analogous to 3.78 that is applicable to finite sets, with unions analogous to sums of subspaces and disjoint unions analogous to direct sums.

20 Suppose U is a subspace of V. Define $\Gamma \colon \mathcal{L}(V/U, W) \to \mathcal{L}(V, W)$ by

$$\Gamma(S) = S \circ \pi.$$

(a) Show that Γ is a linear map.

(b) Show that Γ is injective.

(c) Show that $\operatorname{range} \Gamma = \{T \in \mathcal{L}(V, W) : Tu = 0 \text{ for every } u \in U\}$.

3.F Duality

The Dual Space and the Dual Map

Linear maps into the scalar field $\mathbf{F}$ play a special role in linear algebra, and thus they get a special name:

3.92 Definition ***linear functional***

A ***linear functional*** on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

3.93 Example ***linear functionals***

- Define $\varphi\colon \mathbf{R}^3 \to \mathbf{R}$ by $\varphi(x, y, z) = 4x - 5y + 2z$. Then φ is a linear functional on $\mathbf{R}^3$.
- Fix $(c_1, \dots, c_n) \in \mathbf{F}^n$. Define $\varphi\colon \mathbf{F}^n \to \mathbf{F}$ by
$$\varphi(x_1, \dots, x_n) = c_1 x_1 + \cdots + c_n x_n.$$
Then φ is a linear functional on $\mathbf{F}^n$.
- Define $\varphi\colon \mathcal{P}(\mathbf{R}) \to \mathbf{R}$ by $\varphi(p) = 3p''(5) + 7p(4)$. Then φ is a linear functional on $\mathcal{P}(\mathbf{R})$.
- Define $\varphi\colon \mathcal{P}(\mathbf{R}) \to \mathbf{R}$ by $\varphi(p) = \int_0^1 p(x)\,dx$. Then φ is a linear functional on $\mathcal{P}(\mathbf{R})$.

The vector space $\mathcal{L}(V, \mathbf{F})$ also gets a special name and special notation:

3.94 Definition ***dual space, V'***

The ***dual space*** of V, denoted V', is the vector space of all linear functionals on V. In other words, $V' = \mathcal{L}(V, \mathbf{F})$.

3.95 $\dim V' = \dim V$

Suppose V is finite-dimensional. Then V' is also finite-dimensional and $\dim V' = \dim V$.

Proof This result follows from 3.61. ■

In the following definition, 3.5 implies that each φ_j is well defined.

3.96 Definition ***dual basis***

If $v_1, \dots, v_n$ is a basis of V, then the ***dual basis*** of $v_1, \dots, v_n$ is the list $\varphi_1, \dots, \varphi_n$ of elements of V', where each φ_j is the linear functional on V such that

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

3.97 Example What is the dual basis of the standard basis $e_1, \dots, e_n$ of $\mathbf{F}^n$?

Solution For $1 \leq j \leq n$, define φ_j to be the linear functional on $\mathbf{F}^n$ that selects the j^{th} coordinate of a vector in $\mathbf{F}^n$. In other words,

$$\varphi_j(x_1, \dots, x_n) = x_j$$

for $(x_1, \dots, x_n) \in \mathbf{F}^n$. Clearly

$$\varphi_j(e_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Thus $\varphi_1, \dots, \varphi_n$ is the dual basis of the standard basis $e_1, \dots, e_n$ of $\mathbf{F}^n$.

The next result shows that the dual basis is indeed a basis. Thus the terminology "dual basis" is justified.

3.98 Dual basis is a basis of the dual space

Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V'.

Proof Suppose $v_1, \dots, v_n$ is a basis of V. Let $\varphi_1, \dots, \varphi_n$ denote the dual basis.

To show that $\varphi_1, \dots, \varphi_n$ is a linearly independent list of elements of V', suppose $a_1, \dots, a_n \in F$ are such that

$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0.$$

Now $(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_j) = a_j$ for $j = 1, \dots, n$. The equation above thus shows that $a_1 = \cdots = a_n = 0$. Hence $\varphi_1, \dots, \varphi_n$ is linearly independent.

Now 2.39 and 3.95 imply that $\varphi_1, \dots, \varphi_n$ is a basis of V'. ■

In the definition below, note that if T is a linear map from V to W then T' is a linear map from W' to V'.

3.99 Definition ***dual map, T'***

If $T \in \mathcal{L}(V, W)$, then the ***dual map*** of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$.

If $T \in \mathcal{L}(V, W)$ and $\varphi \in W'$, then $T'(\varphi)$ is defined above to be the composition of the linear maps φ and T. Thus $T'(\varphi)$ is indeed a linear map from V to $\mathbf{F}$; in other words, $T'(\varphi) \in V'$.

The verification that T' is a linear map from W' to V' is easy:

- If $\varphi, \psi \in W'$, then
$$T'(\varphi + \psi) = (\varphi + \psi) \circ T = \varphi \circ T + \psi \circ T = T'(\varphi) + T'(\psi).$$
- If $\lambda \in \mathbf{F}$ and $\varphi \in W'$, then
$$T'(\lambda\varphi) = (\lambda\varphi) \circ T = \lambda(\varphi \circ T) = \lambda T'(\varphi).$$

In the next example, the prime notation is used with two unrelated meanings: D' denotes the dual of a linear map D, and p' denotes the derivative of a polynomial p.

3.100 Example Define $D : \mathcal{P}(\mathbf{R}) \to \mathcal{P}(\mathbf{R})$ by $Dp = p'$.

- Suppose φ is the linear functional on $\mathcal{P}(\mathbf{R})$ defined by $\varphi(p) = p(3)$. Then $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ given by
$$\big(D'(\varphi)\big)(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = p'(3).$$
In other words, $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ that takes p to $p'(3)$.
- Suppose φ is the linear functional on $\mathcal{P}(\mathbf{R})$ defined by $\varphi(p) = \int_0^1 p$. Then $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ given by
$$\big(D'(\varphi)\big)(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = \int_0^1 p' = p(1) - p(0).$$
In other words, $D'(\varphi)$ is the linear functional on $\mathcal{P}(\mathbf{R})$ that takes p to $p(1) - p(0)$.

The first two bullet points in the result below imply that the function that takes T to T' is a linear map from $\mathcal{L}(V, W)$ to $\mathcal{L}(W', V')$.

In the third bullet point below, note the reversal of order from ST on the left to $T'S'$ on the right (here we assume that U is a vector space over $\mathbf{F}$).

3.101 Algebraic properties of dual maps

- $(S+T)' = S' + T'$ for all $S, T \in \mathcal{L}(V, W)$.
- $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbf{F}$ and all $T \in \mathcal{L}(V, W)$.
- $(ST)' = T'S'$ for all $T \in \mathcal{L}(U, V)$ and all $S \in \mathcal{L}(V, W)$.

Proof The proofs of the first two bullet points above are left to the reader.

To prove the third bullet point, suppose $\varphi \in W'$. Then

$$(ST)'(\varphi) = \varphi \circ (ST) = (\varphi \circ S) \circ T = T'(\varphi \circ S) = T'\big(S'(\varphi)\big) = (T'S')(\varphi),$$

where the first, third, and fourth equalities above hold because of the definition of the dual map, the second equality holds because composition of functions is associative, and the last equality follows from the definition of composition.

The equality of the first and last terms above for all $\varphi \in W'$ means that $(ST)' = T'S'$. ■

Some books use the notation V^* and T^* for duality instead of V' and T'. However, here we reserve the notation T^* for the adjoint, which will be introduced when we study linear maps on inner product spaces in Chapter 7.

The Null Space and Range of the Dual of a Linear Map

Our goal in this subsection is to describe null T' and range T' in terms of range T and null T. To do this, we will need the following definition.

3.102 Definition ***annihilator, U^0***

For $U \subset V$, the ***annihilator*** of U, denoted U^0, is defined by

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}.$$

3.103 Example Suppose U is the subspace of $\mathcal{P}(\mathbf{R})$ consisting of all polynomial multiples of x^2. If φ is the linear functional on $\mathcal{P}(\mathbf{R})$ defined by $\varphi(p) = p'(0)$, then $\varphi \in U^0$.

For $U \subset V$, the annihilator U^0 is a subset of the dual space V'. Thus U^0 depends on the vector space containing U, so a notation such as U_V^0 would be more precise. However, the containing vector space will always be clear from the context, so we will use the simpler notation U^0.

3.104 Example Let e_1, e_2, e_3, e_4, e_5 denote the standard basis of $\mathbf{R}^5$, and let $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5$ denote the dual basis of $(\mathbf{R}^5)'$. Suppose

$$U = \operatorname{span}(e_1, e_2) = \{(x_1, x_2, 0, 0, 0) \in \mathbf{R}^5 : x_1, x_2 \in \mathbf{R}\}.$$

Show that $U^0 = \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$.

Solution Recall (see 3.97) that φ_j is the linear functional on $\mathbf{R}^5$ that selects that j^{th} coordinate: $\varphi_j(x_1, x_2, x_3, x_4, x_5) = x_j$.

First suppose $\varphi \in \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$. Then there exist $c_3, c_4, c_5 \in \mathbf{R}$ such that $\varphi = c_3\varphi_3 + c_4\varphi_4 + c_5\varphi_5$. If $(x_1, x_2, 0, 0, 0) \in U$, then

$$\varphi(x_1, x_2, 0, 0, 0) = (c_3\varphi_3 + c_4\varphi_4 + c_5\varphi_5)(x_1, x_2, 0, 0, 0) = 0.$$

Thus $\varphi \in U^0$. In other words, we have shown that $\operatorname{span}(\varphi_3, \varphi_4, \varphi_5) \subset U^0$.

To show the inclusion in the other direction, suppose $\varphi \in U^0$. Because the dual basis is a basis of $(\mathbf{R}^5)'$, there exist $c_1, c_2, c_3, c_4, c_5 \in \mathbf{R}$ such that $\varphi = c_1\varphi_1 + c_2\varphi_2 + c_3\varphi_3 + c_4\varphi_4 + c_5\varphi_5$. Because $e_1 \in U$ and $\varphi \in U^0$, we have

$$0 = \varphi(e_1) = (c_1\varphi_1 + c_2\varphi_2 + c_3\varphi_3 + c_4\varphi_4 + c_5\varphi_5)(e_1) = c_1.$$

Similarly, $e_2 \in U$ and thus $c_2 = 0$. Hence $\varphi = c_3\varphi_3 + c_4\varphi_4 + c_5\varphi_5$. Thus $\varphi \in \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$, which shows that $U^0 \subset \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$.

3.105 The annihilator is a subspace

Suppose $U \subset V$. Then U^0 is a subspace of V'.

Proof Clearly $0 \in U^0$ (here 0 is the zero linear functional on V), because the zero linear functional applied to every vector in U is 0.

Suppose $\varphi, \psi \in U^0$. Thus $\varphi, \psi \in V'$ and $\varphi(u) = \psi(u) = 0$ for every $u \in U$. If $u \in U$, then $(\varphi + \psi)(u) = \varphi(u) + \psi(u) = 0 + 0 = 0$. Thus $\varphi + \psi \in U^0$.

Similarly, U^0 is closed under scalar multiplication. Thus 1.34 implies that U^0 is a subspace of V'. ■

The next result shows that $\dim U^0$ is the difference of $\dim V$ and $\dim U$. For example, this shows that if U is a 2-dimensional subspace of $\mathbf{R}^5$, then U^0 is a 3-dimensional subspace of $(\mathbf{R}^5)'$, as in Example 3.104.

The next result can be proved following the pattern of Example 3.104: choose a basis $u_1, \dots, u_m$ of U, extend to a basis $u_1, \dots, u_m, \dots, u_n$ of V, let $\varphi_1, \dots, \varphi_m, \dots, \varphi_n$ be the dual basis of V', and then show $\varphi_{m+1}, \dots, \varphi_n$ is a basis of U^0, which implies the desired result.

You should construct the proof outlined in the paragraph above, even though a slicker proof is presented here.

3.106 Dimension of the annihilator

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^0 = \dim V.$$

Proof Let $i \in \mathcal{L}(U, V)$ be the inclusion map defined by $i(u) = u$ for $u \in U$. Thus i' is a linear map from V' to U'. The Fundamental Theorem of Linear Maps (3.22) applied to i' shows that

$$\dim \operatorname{range} i' + \dim \operatorname{null} i' = \dim V'.$$

However, $\operatorname{null} i' = U^0$ (as can be seen by thinking about the definitions) and $\dim V' = \dim V$ (by 3.95), so we can rewrite the equation above as

$$\dim \operatorname{range} i' + \dim U^0 = \dim V.$$

If $\varphi \in U'$, then φ can be extended to a linear functional ψ on V (see, for example, Exercise 11 in Section 3.A). The definition of i' shows that $i'(\psi) = \varphi$. Thus $\varphi \in \operatorname{range} i'$, which implies that $\operatorname{range} i' = U'$. Hence $\dim \operatorname{range} i' = \dim U' = \dim U$, and the displayed equation above becomes the desired result. ■

The proof of part (a) of the result below does not use the hypothesis that V and W are finite-dimensional.

3.107 The null space of T'

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

(a) $\operatorname{null} T' = (\operatorname{range} T)^0$;

(b) $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$.

Proof

(a) First suppose $\varphi \in \operatorname{null} T'$. Thus $0 = T'(\varphi) = \varphi \circ T$. Hence

$$0 = (\varphi \circ T)(v) = \varphi(Tv) \quad \text{for every } v \in V.$$

Thus $\varphi \in (\operatorname{range} T)^0$. This implies that $\operatorname{null} T' \subset (\operatorname{range} T)^0$.

To prove the inclusion in the opposite direction, now suppose that $\varphi \in (\operatorname{range} T)^0$. Thus $\varphi(Tv) = 0$ for every vector $v \in V$. Hence $0 = \varphi \circ T = T'(\varphi)$. In other words, $\varphi \in \operatorname{null} T'$, which shows that $(\operatorname{range} T)^0 \subset \operatorname{null} T'$, completing the proof of (a).

(b) We have

$$\begin{aligned} \dim \operatorname{null} T' &= \dim(\operatorname{range} T)^0 \\ &= \dim W - \dim \operatorname{range} T \\ &= \dim W - (\dim V - \dim \operatorname{null} T) \\ &= \dim \operatorname{null} T + \dim W - \dim V, \end{aligned}$$

where the first equality comes from (a), the second equality comes from 3.106, and the third equality comes from the Fundamental Theorem of Linear Maps (3.22). ■

The next result can be useful because sometimes it is easier to verify that T' is injective than to show directly that T is surjective.

3.108 T surjective is equivalent to T' injective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then T is surjective if and only if T' is injective.

Proof The map $T \in \mathcal{L}(V, W)$ is surjective if and only if $\operatorname{range} T = W$, which happens if and only if $(\operatorname{range} T)^0 = \{0\}$, which happens if and only if $\operatorname{null} T' = \{0\}$ [by 3.107(a)], which happens if and only if T' is injective. ■

3.109 The range of T'

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then

(a) $\dim \operatorname{range} T' = \dim \operatorname{range} T$;

(b) $\operatorname{range} T' = (\operatorname{null} T)^0$.

Proof

(a) We have

$$\begin{aligned}\dim \operatorname{range} T' &= \dim W' - \dim \operatorname{null} T' \\ &= \dim W - \dim(\operatorname{range} T)^0 \\ &= \dim \operatorname{range} T,\end{aligned}$$

where the first equality comes from the Fundamental Theorem of Linear Maps (3.22), the second equality comes from 3.95 and 3.107(a), and the third equality comes from 3.106.

(b) First suppose $\varphi \in \operatorname{range} T'$. Thus there exists $\psi \in W'$ such that $\varphi = T'(\psi)$. If $v \in \operatorname{null} T$, then

$$\varphi(v) = \big(T'(\psi)\big)v = (\psi \circ T)(v) = \psi(Tv) = \psi(0) = 0.$$

Hence $\varphi \in (\operatorname{null} T)^0$. This implies that $\operatorname{range} T' \subset (\operatorname{null} T)^0$.

We will complete the proof by showing that $\operatorname{range} T'$ and $(\operatorname{null} T)^0$ have the same dimension. To do this, note that

$$\begin{aligned}\dim \operatorname{range} T' &= \dim \operatorname{range} T \\ &= \dim V - \dim \operatorname{null} T \\ &= \dim(\operatorname{null} T)^0,\end{aligned}$$

where the first equality comes from (a), the second equality comes from the Fundamental Theorem of Linear Maps (3.22), and the third equality comes from 3.106. ■

The next result should be compared to 3.108.

3.110 T injective is equivalent to T' surjective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then T is injective if and only if T' is surjective.

Proof The map $T \in \mathcal{L}(V, W)$ is injective if and only if $\operatorname{null} T = \{0\}$, which happens if and only if $(\operatorname{null} T)^0 = V'$, which happens if and only if $\operatorname{range} T' = V'$ [by 3.109(b)], which happens if and only if T' is surjective. ■

The Matrix of the Dual of a Linear Map

We now define the transpose of a matrix.

3.111 Definition ***transpose,*** A^{t}

The ***transpose*** of a matrix A, denoted A^{t}, is the matrix obtained from A by interchanging the rows and columns. More specifically, if A is an m-by-n matrix, then A^{t} is the n-by-m matrix whose entries are given by the equation

$$(A^{\mathrm{t}})_{k,j} = A_{j,k}.$$

3.112 Example If $A = \begin{pmatrix} 5 & -7 \\ 3 & 8 \\ -4 & 2 \end{pmatrix}$, then $A^{\mathrm{t}} = \begin{pmatrix} 5 & 3 & -4 \\ -7 & 8 & 2 \end{pmatrix}$.

Note that here A is a 3-by-2 matrix and A^{t} is a 2-by-3 matrix.

The transpose has nice algebraic properties: $(A + C)^{\mathrm{t}} = A^{\mathrm{t}} + C^{\mathrm{t}}$ and $(\lambda A)^{\mathrm{t}} = \lambda A^{\mathrm{t}}$ for all m-by-n matrices A, C and all $\lambda \in \mathbf{F}$ (see Exercise 33).

The next result shows that the transpose of the product of two matrices is the product of the transposes in the opposite order.

3.113 The transpose of the product of matrices

If A is an m-by-n matrix and C is an n-by-p matrix, then

$$(AC)^{\mathrm{t}} = C^{\mathrm{t}} A^{\mathrm{t}}.$$

Proof Suppose $1 \leq k \leq p$ and $1 \leq j \leq m$. Then

$$\begin{aligned} \big((AC)^{\mathrm{t}}\big)_{k,j} &= (AC)_{j,k} \\ &= \sum_{r=1}^{n} A_{j,r} C_{r,k} \\ &= \sum_{r=1}^{n} (C^{\mathrm{t}})_{k,r} (A^{\mathrm{t}})_{r,j} \\ &= (C^{\mathrm{t}} A^{\mathrm{t}})_{k,j}. \end{aligned}$$

Thus $(AC)^{\mathrm{t}} = C^{\mathrm{t}} A^{\mathrm{t}}$, as desired. ■

The setting for the next result is the assumption that we have a basis $v_1, \ldots, v_n$ of V, along with its dual basis $\varphi_1, \ldots, \varphi_n$ of V'. We also have a basis $w_1, \ldots, w_m$ of W, along with its dual basis $\psi_1, \ldots, \psi_m$ of W'. Thus $\mathcal{M}(T)$ is computed with respect to the bases just mentioned of V and W, and $\mathcal{M}(T')$ is computed with respect to the dual bases just mentioned of W' and V'.

3.114 The matrix of T' is the transpose of the matrix of T

Suppose $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(T') = \big(\mathcal{M}(T)\big)^{\mathrm{t}}$.

Proof Let $A = \mathcal{M}(T)$ and $C = \mathcal{M}(T')$. Suppose $1 \leq j \leq m$ and $1 \leq k \leq n$.

From the definition of $\mathcal{M}(T')$ we have

$$T'(\psi_j) = \sum_{r=1}^{n} C_{r,j}\varphi_r.$$

The left side of the equation above equals $\psi_j \circ T$. Thus applying both sides of the equation above to v_k gives

$$\begin{aligned}(\psi_j \circ T)(v_k) &= \sum_{r=1}^{n} C_{r,j}\varphi_r(v_k)\\ &= C_{k,j}.\end{aligned}$$

We also have

$$\begin{aligned}(\psi_j \circ T)(v_k) &= \psi_j(Tv_k)\\ &= \psi_j\Big(\sum_{r=1}^{m} A_{r,k}w_r\Big)\\ &= \sum_{r=1}^{m} A_{r,k}\psi_j(w_r)\\ &= A_{j,k}.\end{aligned}$$

Comparing the last line of the last two sets of equations, we have $C_{k,j} = A_{j,k}$. Thus $C = A^{\mathrm{t}}$. In other words, $\mathcal{M}(T') = \big(\mathcal{M}(T)\big)^{\mathrm{t}}$, as desired. ■

The Rank of a Matrix

We begin by defining two nonnegative integers that are associated with each matrix.

3.115 Definition ***row rank, column rank***

Suppose A is an m-by-n matrix with entries in $\mathbf{F}$.

- The ***row rank*** of A is the dimension of the span of the rows of A in $\mathbf{F}^{1,n}$.
- The ***column rank*** of A is the dimension of the span of the columns of A in $\mathbf{F}^{m,1}$.

3.116 Example Suppose $A = \begin{pmatrix} 4 & 7 & 1 & 8 \\ 3 & 5 & 2 & 9 \end{pmatrix}$. Find the row rank of A and the column rank of A.

Solution The row rank of A is the dimension of

$$\operatorname{span}\left(\begin{pmatrix} 4 & 7 & 1 & 8 \end{pmatrix}, \begin{pmatrix} 3 & 5 & 2 & 9 \end{pmatrix}\right)$$

in $\mathbf{F}^{1,4}$. Neither of the two vectors listed above in $\mathbf{F}^{1,4}$ is a scalar multiple of the other. Thus the span of this list of length 2 has dimension 2. In other words, the row rank of A is 2.

The column rank of A is the dimension of

$$\operatorname{span}\left(\begin{pmatrix} 4 \\ 3 \end{pmatrix}, \begin{pmatrix} 7 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 8 \\ 9 \end{pmatrix}\right)$$

in $\mathbf{F}^{2,1}$. Neither of the first two vectors listed above in $\mathbf{F}^{2,1}$ is a scalar multiple of the other. Thus the span of this list of length 4 has dimension at least 2. The span of this list of vectors in $\mathbf{F}^{2,1}$ cannot have dimension larger than 2 because $\dim \mathbf{F}^{2,1} = 2$. Thus the span of this list has dimension 2. In other words, the column rank of A is 2.

Notice that no bases are in sight in the statement of the next result. Although $\mathcal{M}(T)$ in the next result depends on a choice of bases of V and W, the next result shows that the column rank of $\mathcal{M}(T)$ is the same for all such choices (because range T does not depend on a choice of basis).

3.117 Dimension of range T equals column rank of $\mathcal{M}(T)$

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\dim \operatorname{range} T$ equals the column rank of $\mathcal{M}(T)$.

Proof Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. The function that takes $w \in \operatorname{span}(Tv_1, \ldots, Tv_n)$ to $\mathcal{M}(w)$ is easily seen to be an isomorphism from $\operatorname{span}(Tv_1, \ldots, Tv_n)$ onto $\operatorname{span}\big(\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)\big)$. Thus $\dim \operatorname{span}(Tv_1, \ldots, Tv_n) = \dim \operatorname{span}\big(\mathcal{M}(Tv_1), \ldots, \mathcal{M}(Tv_n)\big)$, where the last dimension equals the column rank of $\mathcal{M}(T)$.

It is easy to see that $\operatorname{range} T = \operatorname{span}(Tv_1, \ldots, Tv_n)$. Thus we have $\dim \operatorname{range} T = \dim \operatorname{span}(Tv_1, \ldots, Tv_n) =$ the column rank of $\mathcal{M}(T)$, as desired. ■

In Example 3.116, the row rank and column rank turned out to equal each other. The next result shows that this always happens.

3.118 Row rank equals column rank

Suppose $A \in \mathbf{F}^{m,n}$. Then the row rank of A equals the column rank of A.

Proof Define $T: \mathbf{F}^{n,1} \to \mathbf{F}^{m,1}$ by $Tx = Ax$. Thus $\mathcal{M}(T) = A$, where $\mathcal{M}(T)$ is computed with respect to the standard bases of $\mathbf{F}^{n,1}$ and $\mathbf{F}^{m,1}$. Now

$$\begin{aligned}
\text{column rank of } A &= \text{column rank of } \mathcal{M}(T) \\
&= \dim \operatorname{range} T \\
&= \dim \operatorname{range} T' \\
&= \text{column rank of } \mathcal{M}(T') \\
&= \text{column rank of } A^{\mathrm{t}} \\
&= \text{row rank of } A,
\end{aligned}$$

where the second equality above comes from 3.117, the third equality comes from 3.109(a), the fourth equality comes from 3.117 (where $\mathcal{M}(T')$ is computed with respect to the dual bases of the standard bases), the fifth equality comes from 3.114, and the last equality follows easily from the definitions. ■

The last result allows us to dispense with the terms "row rank" and "column rank" and just use the simpler term "rank".

3.119 Definition *rank*

The ***rank*** of a matrix $A \in \mathbf{F}^{m,n}$ is the column rank of A.

EXERCISES 3.F

1 Explain why every linear functional is either surjective or the zero map.

2 Give three distinct examples of linear functionals on $\mathbf{R}^{[0,1]}$.

3 Suppose V is finite-dimensional and $v \in V$ with $v \neq 0$. Prove that there exists $\varphi \in V'$ such that $\varphi(v) = 1$.

4 Suppose V is finite-dimensional and U is a subspace of V such that $U \neq V$. Prove that there exists $\varphi \in V'$ such that $\varphi(u) = 0$ for every $u \in U$ but $\varphi \neq 0$.

5 Suppose $V_1, \dots, V_m$ are vector spaces. Prove that $(V_1 \times \cdots \times V_m)'$ and $V_1{}' \times \cdots \times V_m{}'$ are isomorphic vector spaces.

6 Suppose V is finite-dimensional and $v_1, \dots, v_m \in V$. Define a linear map $\Gamma\colon V' \to \mathbf{F}^m$ by

$$\Gamma(\varphi) = \big(\varphi(v_1), \dots, \varphi(v_m)\big).$$

(a) Prove that $v_1, \dots, v_m$ spans V if and only if Γ is injective.

(b) Prove that $v_1, \dots, v_m$ is linearly independent if and only if Γ is surjective.

7 Suppose m is a positive integer. Show that the dual basis of the basis $1, x, \dots, x^m$ of $\mathcal{P}_m(\mathbf{R})$ is $\varphi_0, \varphi_1, \dots, \varphi_m$, where $\varphi_j(p) = \frac{p^{(j)}(0)}{j!}$. Here $p^{(j)}$ denotes the j^{th} derivative of p, with the understanding that the 0^{th} derivative of p is p.

8 Suppose m is a positive integer.

(a) Show that $1, x-5, \dots, (x-5)^m$ is a basis of $\mathcal{P}_m(\mathbf{R})$.

(b) What is the dual basis of the basis in part (a)?

9 Suppose $v_1, \dots, v_n$ is a basis of V and $\varphi_1, \dots, \varphi_n$ is the corresponding dual basis of V'. Suppose $\psi \in V'$. Prove that

$$\psi = \psi(v_1)\varphi_1 + \cdots + \psi(v_n)\varphi_n.$$

10 Prove the first two bullet points in 3.101.

11 Suppose A is an m-by-n matrix with $A \neq 0$. Prove that the rank of A is 1 if and only if there exist $(c_1, \dots, c_m) \in \mathbf{F}^m$ and $(d_1, \dots, d_n) \in \mathbf{F}^n$ such that $A_{j,k} = c_j d_k$ for every $j = 1, \dots, m$ and every $k = 1, \dots, n$.

12 Show that the dual map of the identity map on V is the identity map on V'.

13 Define $T : \mathbf{R}^3 \to \mathbf{R}^2$ by $T(x, y, z) = (4x + 5y + 6z, 7x + 8y + 9z)$. Suppose φ_1, φ_2 denotes the dual basis of the standard basis of $\mathbf{R}^2$ and ψ_1, ψ_2, ψ_3 denotes the dual basis of the standard basis of $\mathbf{R}^3$.

(a) Describe the linear functionals $T'(\varphi_1)$ and $T'(\varphi_2)$.

(b) Write $T'(\varphi_1)$ and $T'(\varphi_2)$ as linear combinations of ψ_1, ψ_2, ψ_3.

14 Define $T : \mathcal{P}(\mathbf{R}) \to \mathcal{P}(\mathbf{R})$ by $(Tp)(x) = x^2 p(x) + p''(x)$ for $x \in \mathbf{R}$.

(a) Suppose $\varphi \in \mathcal{P}(\mathbf{R})'$ is defined by $\varphi(p) = p'(4)$. Describe the linear functional $T'(\varphi)$ on $\mathcal{P}(\mathbf{R})$.

(b) Suppose $\varphi \in \mathcal{P}(\mathbf{R})'$ is defined by $\varphi(p) = \int_0^1 p(x)\,dx$. Evaluate $\big(T'(\varphi)\big)(x^3)$.

15 Suppose W is finite-dimensional and $T \in \mathcal{L}(V, W)$. Prove that $T' = 0$ if and only if $T = 0$.

16 Suppose V and W are finite-dimensional. Prove that the map that takes $T \in \mathcal{L}(V, W)$ to $T' \in \mathcal{L}(W', V')$ is an isomorphism of $\mathcal{L}(V, W)$ onto $\mathcal{L}(W', V')$.

17 Suppose $U \subset V$. Explain why $U^0 = \{\varphi \in V' : U \subset \operatorname{null} \varphi\}$.

18 Suppose V is finite-dimensional and $U \subset V$. Show that $U = \{0\}$ if and only if $U^0 = V'$.

19 Suppose V is finite-dimensional and U is a subspace of V. Show that $U = V$ if and only if $U^0 = \{0\}$.

20 Suppose U and W are subsets of V with $U \subset W$. Prove that $W^0 \subset U^0$.

21 Suppose V is finite-dimensional and U and W are subspaces of V with $W^0 \subset U^0$. Prove that $U \subset W$.

22 Suppose U, W are subspaces of V. Show that $(U + W)^0 = U^0 \cap W^0$.

23 Suppose V is finite-dimensional and U and W are subspaces of V. Prove that $(U \cap W)^0 = U^0 + W^0$.

24 Prove 3.106 using the ideas sketched in the discussion before the statement of 3.106.

25 Suppose V is finite-dimensional and U is a subspace of V. Show that

$$U = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in U^0\}.$$

26 Suppose V is finite-dimensional and Γ is a subspace of V'. Show that

$$\Gamma = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in \Gamma\}^0.$$

27 Suppose $T \in \mathcal{L}\big(\mathcal{P}_5(\mathbf{R}), \mathcal{P}_5(\mathbf{R})\big)$ and $\operatorname{null} T' = \operatorname{span}(\varphi)$, where φ is the linear functional on $\mathcal{P}_5(\mathbf{R})$ defined by $\varphi(p) = p(8)$. Prove that $\operatorname{range} T = \{p \in \mathcal{P}_5(\mathbf{R}) : p(8) = 0\}$.

28 Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V, W)$, and there exists $\varphi \in W'$ such that $\operatorname{null} T' = \operatorname{span}(\varphi)$. Prove that $\operatorname{range} T = \operatorname{null} \varphi$.

29 Suppose V and W are finite-dimensional, $T \in \mathcal{L}(V, W)$, and there exists $\varphi \in V'$ such that $\operatorname{range} T' = \operatorname{span}(\varphi)$. Prove that $\operatorname{null} T = \operatorname{null} \varphi$.

30 Suppose V is finite-dimensional and $\varphi_1, \dots, \varphi_m$ is a linearly independent list in V'. Prove that

$$\dim\big((\operatorname{null} \varphi_1) \cap \cdots \cap (\operatorname{null} \varphi_m)\big) = (\dim V) - m.$$

31 Suppose V is finite-dimensional and $\varphi_1, \dots, \varphi_n$ is a basis of V'. Show that there exists a basis of V whose dual basis is $\varphi_1, \dots, \varphi_n$.

32 Suppose $T \in \mathcal{L}(V)$, and $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Prove that the following are equivalent:

(a) T is invertible.

(b) The columns of $\mathcal{M}(T)$ are linearly independent in $\mathbf{F}^{n,1}$.

(c) The columns of $\mathcal{M}(T)$ span $\mathbf{F}^{n,1}$.

(d) The rows of $\mathcal{M}(T)$ are linearly independent in $\mathbf{F}^{1,n}$.

(e) The rows of $\mathcal{M}(T)$ span $\mathbf{F}^{1,n}$.

Here $\mathcal{M}(T)$ means $\mathcal{M}\big(T, (u_1, \dots, u_n), (v_1, \dots, u_n)\big)$.

33 Suppose m and n are positive integers. Prove that the function that takes A to A^{t} is a linear map from $\mathbf{F}^{m,n}$ to $\mathbf{F}^{n,m}$. Furthermore, prove that this linear map is invertible.

34 The ***double dual space*** of V, denoted V'', is defined to be the dual space of V'. In other words, $V'' = (V')'$. Define $\Lambda\colon V \to V''$ by

$$(\Lambda v)(\varphi) = \varphi(v)$$

for $v \in V$ and $\varphi \in V'$.

(a) Show that Λ is a linear map from V to V''.

(b) Show that if $T \in \mathcal{L}(V)$, then $T'' \circ \Lambda = \Lambda \circ T$, where $T'' = (T')'$.

(c) Show that if V is finite-dimensional, then Λ is an isomorphism from V onto V''.

[*Suppose V is finite-dimensional. Then V and V' are isomorphic, but finding an isomorphism from V onto V' generally requires choosing a basis of V. In contrast, the isomorphism Λ from V onto V'' does not require a choice of basis and thus is considered more natural.*]

35 Show that $\big(\mathcal{P}(\mathbf{R})\big)'$ and $\mathbf{R}^\infty$ are isomorphic.

36 Suppose U is a subspace of V. Let $i\colon U \to V$ be the inclusion map defined by $i(u) = u$. Thus $i' \in \mathcal{L}(V', U')$.

(a) Show that $\operatorname{null} i' = U^0$.

(b) Prove that if V is finite-dimensional, then $\operatorname{range} i' = U'$.

(c) Prove that if V is finite-dimensional, then $\widetilde{i'}$ is an isomorphism from V'/U^0 onto U'.

[*The isomorphism in part (c) is natural in that it does not depend on a choice of basis in either vector space.*]

37 Suppose U is a subspace of V. Let $\pi\colon V \to V/U$ be the usual quotient map. Thus $\pi' \in \mathcal{L}\big((V/U)', V'\big)$.

(a) Show that π' is injective.

(b) Show that $\operatorname{range} \pi' = U^0$.

(c) Conclude that π' is an isomorphism from $(V/U)'$ onto U^0.

[*The isomorphism in part (c) is natural in that it does not depend on a choice of basis in either vector space. In fact, there is no assumption here that any of these vector spaces are finite-dimensional.*]

CHAPTER

4

Statue of Persian mathematician and poet Omar Khayyám (1048–1131), whose algebra book written in 1070 contained the first serious study of cubic polynomials.

Polynomials

This short chapter contains material on polynomials that we will need to understand operators. Many of the results in this chapter will already be familiar to you from other courses; they are included here for completeness.

Because this chapter is not about linear algebra, your instructor may go through it rapidly. You may not be asked to scrutinize all the proofs. Make sure, however, that you at least read and understand the statements of all the results in this chapter—they will be used in later chapters.

The standing assumption we need for this chapter is as follows:

4.1 Notation $\mathbf{F}$

$\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- Division Algorithm for Polynomials
- factorization of polynomials over $\mathbf{C}$
- factorization of polynomials over $\mathbf{R}$

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_4

Complex Conjugate and Absolute Value

Before discussing polynomials with complex or real coefficients, we need to learn a bit more about the complex numbers.

4.2 Definition Re z, Im z

Suppose $z = a + bi$, where a and b are real numbers.

- The ***real part*** of z, denoted $\operatorname{Re} z$, is defined by $\operatorname{Re} z = a$.
- The ***imaginary part*** of z, denoted $\operatorname{Im} z$, is defined by $\operatorname{Im} z = b$.

Thus for every complex number z, we have

$$z = \operatorname{Re} z + (\operatorname{Im} z)i.$$

4.3 Definition ***complex conjugate, $\bar{z}$, absolute value, $|z|$***

Suppose $z \in \mathbf{C}$.

- The ***complex conjugate*** of $z \in \mathbf{C}$, denoted $\bar{z}$, is defined by
$$\bar{z} = \operatorname{Re} z - (\operatorname{Im} z)i.$$
- The ***absolute value*** of a complex number z, denoted $|z|$, is defined by
$$|z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

4.4 Example Suppose $z = 3 + 2i$. Then

- $\operatorname{Re} z = 3$ and $\operatorname{Im} z = 2$;
- $\bar{z} = 3 - 2i$;
- $|z| = \sqrt{3^2 + 2^2} = \sqrt{13}$.

Note that $|z|$ is a nonnegative number for every $z \in \mathbf{C}$.

You should verify that $z = \bar{z}$ if and only if z is a real number.

The real and imaginary parts, complex conjugate, and absolute value have the following properties:

4.5 Properties of complex numbers

Suppose $w, z \in \mathbf{C}$. Then

sum of z and $\bar{z}$

$$z + \bar{z} = 2 \operatorname{Re} z;$$

difference of z and $\bar{z}$

$$z - \bar{z} = 2(\operatorname{Im} z)i;$$

product of z and $\bar{z}$

$$z\bar{z} = |z|^2;$$

additivity and multiplicativity of complex conjugate

$$\overline{w + z} = \bar{w} + \bar{z} \text{ and } \overline{wz} = \bar{w}\bar{z};$$

conjugate of conjugate

$$\bar{\bar{z}} = z;$$

real and imaginary parts are bounded by $|z|$

$$|\operatorname{Re} z| \le |z| \text{ and } |\operatorname{Im} z| \le |z|$$

absolute value of the complex conjugate

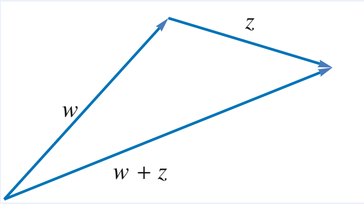
$$|\bar{z}| = |z|;$$

multiplicativity of absolute value

$$|wz| = |w|\,|z|;$$

Triangle Inequality

$$|w + z| \le |w| + |z|.$$



Proof Except for the last item, the routine verifications of the assertions above are left to the reader. To verify the last item, we have

$$\begin{aligned}
|w + z|^2 &= (w + z)(\bar{w} + \bar{z}) \\
&= w\bar{w} + z\bar{z} + w\bar{z} + z\bar{w} \\
&= |w|^2 + |z|^2 + w\bar{z} + \overline{w\bar{z}} \\
&= |w|^2 + |z|^2 + 2 \operatorname{Re}(w\bar{z}) \\
&\le |w|^2 + |z|^2 + 2|w\bar{z}| \\
&= |w|^2 + |z|^2 + 2|w|\,|z| \\
&= (|w| + |z|)^2.
\end{aligned}$$

Taking the square root of both sides of the inequality $|w + z|^2 \le (|w| + |z|)^2$ now gives the desired inequality. ∎

Uniqueness of Coefficients for Polynomials

Recall that a function $p\colon \mathbf{F} \to \mathbf{F}$ is called a polynomial with coefficients in $\mathbf{F}$ if there exist $a_0, \dots, a_m \in \mathbf{F}$ such that

4.6
$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbf{F}$.

4.7 If a polynomial is the zero function, then all coefficients are 0

Suppose $a_0, \dots, a_m \in \mathbf{F}$. If

$$a_0 + a_1 z + \cdots + a_m z^m = 0$$

for every $z \in \mathbf{F}$, then $a_0 = \cdots = a_m = 0$.

Proof We will prove the contrapositive. If not all the coefficients are 0, then by changing m we can assume $a_m \neq 0$. Let

$$z = \frac{|a_0| + |a_1| + \cdots + |a_{m-1}|}{|a_m|} + 1.$$

Note that $z \geq 1$, and thus $z^j \leq z^{m-1}$ for $j = 0, 1, \dots, m-1$. Using the Triangle Inequality, we have

$$\begin{aligned} |a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}| &\leq (|a_0| + |a_1| + \cdots + |a_{m-1}|) z^{m-1} \\ &< |a_m z^m|. \end{aligned}$$

Thus $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} \neq -a_m z^m$. Hence we conclude that $a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + a_m z^m \neq 0$. ■

The result above implies that the coefficients of a polynomial are uniquely determined (because if a polynomial had two different sets of coefficients, then subtracting the two representations of the polynomial would give a contradiction to the result above).

Recall that if a polynomial p can be written in the form 4.6 with $a_m \neq 0$, then we say that p has degree m and we write $\deg p = m$.

The 0 polynomial is declared to have degree $-\infty$ so that exceptions are not needed for various reasonable results. For example, $\deg(pq) = \deg p + \deg q$ even if $p = 0$.

The degree of the 0 polynomial is defined to be $-\infty$. When necessary, use the obvious arithmetic with $-\infty$. For example, $-\infty < m$ and $-\infty + m = -\infty$ for every integer m.

The Division Algorithm for Polynomials

If p and s are nonnegative integers, with $s \neq 0$, then there exist nonnegative integers q and r such that

$$p = sq + r$$

and $r < s$. Think of dividing p by s, getting quotient q with remainder r. Our next task is to prove an analogous result for polynomials.

The result below is often called the Division Algorithm for Polynomials, although as stated here it is not really an algorithm, just a useful result.

Think of the Division Algorithm for Polynomials as giving the remainder r when p is divided by s.

Recall that $\mathcal{P}(\mathbf{F})$ denotes the vector space of all polynomials with coefficients in $\mathbf{F}$ and that $\mathcal{P}_m(\mathbf{F})$ is the subspace of $\mathcal{P}(\mathbf{F})$ consisting of the polynomials with coefficients in $\mathbf{F}$ and degree at most m.

The next result can be proved without linear algebra, but the proof given here using linear algebra is appropriate for a linear algebra textbook.

4.8 Division Algorithm for Polynomials

Suppose that $p, s \in \mathcal{P}(\mathbf{F})$, with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathcal{P}(\mathbf{F})$ such that

$$p = sq + r$$

and $\deg r < \deg s$.

Proof Let $n = \deg p$ and $m = \deg s$. If $n < m$, then take $q = 0$ and $r = p$ to get the desired result. Thus we can assume that $n \geq m$.

Define $T : \mathcal{P}_{n-m}(\mathbf{F}) \times \mathcal{P}_{m-1}(\mathbf{F}) \to \mathcal{P}_n(\mathbf{F})$ by

$$T(q, r) = sq + r.$$

The reader can easily verify that T is a linear map. If $(q, r) \in \operatorname{null} T$, then $sq + r = 0$, which implies that $q = 0$ and $r = 0$ [because otherwise $\deg sq \geq m$ and thus sq cannot equal $-r$]. Thus $\dim \operatorname{null} T = 0$ (proving the "unique" part of the result).

From 3.76 we have

$$\dim\big(\mathcal{P}_{n-m}(\mathbf{F}) \times \mathcal{P}_{m-1}(\mathbf{F})\big) = (n - m + 1) + (m - 1 + 1) = n + 1.$$

The Fundamental Theorem of Linear Maps (3.22) and the equation displayed above now imply that $\dim \operatorname{range} T = n + 1$, which equals $\dim \mathcal{P}_n(\mathbf{F})$. Thus $\operatorname{range} T = \mathcal{P}_n(\mathbf{F})$, and hence there exist $q \in \mathcal{P}_{n-m}(\mathbf{F})$ and $r \in \mathcal{P}_{m-1}(\mathbf{F})$ such that $p = T(q, r) = sq + r$. ■

Zeros of Polynomials

The solutions to the equation $p(z) = 0$ play a crucial role in the study of a polynomial $p \in \mathcal{P}(\mathbf{F})$. Thus these solutions have a special name.

4.9 Definition ***zero of a polynomial***

A number $\lambda \in \mathbf{F}$ is called a ***zero*** (or ***root***) of a polynomial $p \in \mathcal{P}(\mathbf{F})$ if

$$p(\lambda) = 0.$$

4.10 Definition ***factor***

A polynomial $s \in \mathcal{P}(\mathbf{F})$ is called a ***factor*** of $p \in \mathcal{P}(\mathbf{F})$ if there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that $p = sq$.

We begin by showing that λ is a zero of a polynomial $p \in \mathcal{P}(\mathbf{F})$ if and only if $z - \lambda$ is a factor of p.

4.11 Each zero of a polynomial corresponds to a degree-1 factor

Suppose $p \in \mathcal{P}(\mathbf{F})$ and $\lambda \in \mathbf{F}$. Then $p(\lambda) = 0$ if and only if there is a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbf{F}$.

Proof One direction is obvious. Namely, suppose there is a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that $p(z) = (z - \lambda)q(z)$ for all $z \in \mathbf{F}$. Then

$$p(\lambda) = (\lambda - \lambda)q(\lambda) = 0,$$

as desired.

To prove the other direction, suppose $p(\lambda) = 0$. The polynomial $z - \lambda$ has degree 1. Because a polynomial with degree less than 1 is a constant function, the Division Algorithm for Polynomials (4.8) implies that there exist a polynomial $q \in \mathcal{P}(\mathbf{F})$ and a number $r \in \mathbf{F}$ such that

$$p(z) = (z - \lambda)q(z) + r$$

for every $z \in \mathbf{F}$. The equation above and the equation $p(\lambda) = 0$ imply that $r = 0$. Thus $p(z) = (z - \lambda)q(z)$ for every $z \in \mathbf{F}$. ∎

Now we can prove that polynomials do not have too many zeros.

4.12 A polynomial has at most as many zeros as its degree

Suppose $p \in \mathcal{P}(\mathbf{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbf{F}$.

Proof If $m = 0$, then $p(z) = a_0 \neq 0$ and so p has no zeros.

If $m = 1$, then $p(z) = a_0 + a_1 z$, with $a_1 \neq 0$, and thus p has exactly one zero, namely, $-a_0/a_1$.

Now suppose $m > 1$. We use induction on m, assuming that every polynomial with degree $m - 1$ has at most $m - 1$ distinct zeros. If p has no zeros in $\mathbf{F}$, then we are done. If p has a zero $\lambda \in \mathbf{F}$, then by 4.11 there is a polynomial q such that

$$p(z) = (z - \lambda)q(z)$$

for all $z \in \mathbf{F}$. Clearly $\deg q = m - 1$. The equation above shows that if $p(z) = 0$, then either $z = \lambda$ or $q(z) = 0$. In other words, the zeros of p consist of λ and the zeros of q. By our induction hypothesis, q has at most $m - 1$ distinct zeros in $\mathbf{F}$. Thus p has at most m distinct zeros in $\mathbf{F}$. ■

Factorization of Polynomials over C

So far we have been handling polynomials with complex coefficients and polynomials with real coefficients simultaneously through our convention that $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$. Now we will see some differences between these two cases. First we treat polynomials with complex coefficients. Then we will use our results about polynomials with complex coefficients to prove corresponding results for polynomials with real coefficients.

The next result, although called the Fundamental Theorem of Algebra, uses analysis its proof. The short proof presented here uses tools from complex analysis. If you have not had a course in complex analysis, this proof will almost certainly be meaningless to you. In that case, just accept the Fundamental Theorem of Algebra as something that we need to use but whose proof requires more advanced tools that you may learn in later courses.

The Fundamental Theorem of Algebra is an existence theorem. Its proof does not lead to a method for finding zeros. The quadratic formula gives the zeros explicitly for polynomials of degree 2. Similar but more complicated formulas exist for polynomials of degree 3 and 4. No such formulas exist for polynomials of degree 5 and above.

4.13 Fundamental Theorem of Algebra

Every nonconstant polynomial with complex coefficients has a zero.

Proof Let p be a nonconstant polynomial with complex coefficients. Suppose p has no zeros. Then $1/p$ is an analytic function on $\mathbf{C}$. Furthermore, $|p(z)| \to \infty$ as $|z| \to \infty$, which implies that $1/p \to 0$ as $|z| \to \infty$. Thus $1/p$ is a bounded analytic function on $\mathbf{C}$. By Liouville's theorem, every such function is constant. But if $1/p$ is constant, then p is constant, contradicting our assumption that p is nonconstant. ■

Although the proof given above is probably the shortest proof of the Fundamental Theorem of Algebra, a web search can lead you to several other proofs that use different techniques. All proofs of the Fundamental Theorem of Algebra need to use some analysis, because the result is not true if $\mathbf{C}$ is replaced, for example, with the set of numbers of the form $c + di$ where c, d are rational numbers.

The cubic formula, which was discovered in the 16th century, is presented below for your amusement only. Do not memorize it.

Suppose

$$p(x) = ax^3 + bx^2 + cx + d,$$

where $a \neq 0$. *Set*

$$u = \frac{9abc - 2b^3 - 27a^2d}{54a^3}$$

and then set

$$v = u^2 + \left(\frac{3ac - b^2}{9a^2}\right)^3.$$

Suppose $v \geq 0$. *Then*

$$-\frac{b}{3a} + \sqrt[3]{u + \sqrt{v}} + \sqrt[3]{u - \sqrt{v}}$$

is a zero of p.

Remarkably, mathematicians have proved that no formula exists for the zeros of polynomials of degree 5 or higher. But computers and calculators can use clever numerical methods to find good approximations to the zeros of any polynomial, even when exact zeros cannot be found.

For example, no one will ever be able to give an exact formula for a zero of the polynomial p defined by

$$p(x) = x^5 - 5x^4 - 6x^3 + 17x^2 + 4x - 7.$$

However, a computer or symbolic calculator can find approximate zeros of this polynomial.

The Fundamental Theorem of Algebra leads to the following factorization result for polynomials with complex coefficients. Note that in this factorization, the numbers $\lambda_1, \dots, \lambda_m$ are precisely the zeros of p, for these are the only values of z for which the right side of the equation in the next result equals 0.

4.14 Factorization of a polynomial over $\mathbf{C}$

If $p \in \mathcal{P}(\mathbf{C})$ is a nonconstant polynomial, then p has a unique factorization (except for the order of the factors) of the form

$$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m),$$

where $c, \lambda_1, \dots, \lambda_m \in \mathbf{C}$.

Proof Let $p \in \mathcal{P}(\mathbf{C})$ and let $m = \deg p$. We will use induction on m. If $m = 1$, then clearly the desired factorization exists and is unique. So assume that $m > 1$ and that the desired factorization exists and is unique for all polynomials of degree $m - 1$.

First we will show that the desired factorization of p exists. By the Fundamental Theorem of Algebra (4.13), p has a zero λ. By 4.11, there is a polynomial q such that

$$p(z) = (z - \lambda)q(z)$$

for all $z \in \mathbf{C}$. Because $\deg q = m - 1$, our induction hypothesis implies that q has the desired factorization, which when plugged into the equation above gives the desired factorization of p.

Now we turn to the question of uniqueness. Clearly c is uniquely determined as the coefficient of z^m in p. So we need only show that except for the order, there is only one way to choose $\lambda_1, \dots, \lambda_m$. If

$$(z - \lambda_1) \cdots (z - \lambda_m) = (z - \tau_1) \cdots (z - \tau_m)$$

for all $z \in \mathbf{C}$, then because the left side of the equation above equals 0 when $z = \lambda_1$, one of the τ's on the right side equals λ_1. Relabeling, we can assume that $\tau_1 = \lambda_1$. Now for $z \neq \lambda_1$, we can divide both sides of the equation above by $z - \lambda_1$, getting

$$(z - \lambda_2) \cdots (z - \lambda_m) = (z - \tau_2) \cdots (z - \tau_m)$$

for all $z \in \mathbf{C}$ except possibly $z = \lambda_1$. Actually the equation above holds for all $z \in \mathbf{C}$, because otherwise by subtracting the right side from the left side we would get a nonzero polynomial that has infinitely many zeros. The equation above and our induction hypothesis imply that except for the order, the λ's are the same as the τ's, completing the proof of uniqueness. ■

Factorization of Polynomials over R

The failure of the Fundamental Theorem of Algebra for **R** *accounts for the differences between operators on real and complex vector spaces, as we will see in later chapters.*

A polynomial with real coefficients may have no real zeros. For example, the polynomial $1 + x^2$ has no real zeros.

To obtain a factorization theorem over **R**, we will use our factorization theorem over **C**. We begin with the following result.

4.15 Polynomials with real coefficients have zeros in pairs

Suppose $p \in \mathcal{P}(\mathbf{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbf{C}$ is a zero of p, then so is $\bar{\lambda}$.

Proof Let

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m,$$

where $a_0, \ldots, a_m$ are real numbers. Suppose $\lambda \in \mathbf{C}$ is a zero of p. Then

$$a_0 + a_1 \lambda + \cdots + a_m \lambda^m = 0.$$

Take the complex conjugate of both sides of this equation, obtaining

$$a_0 + a_1 \bar{\lambda} + \cdots + a_m \bar{\lambda}^m = 0,$$

where we have used basic properties of complex conjugation (see 4.5). The equation above shows that $\bar{\lambda}$ is a zero of p. ■

Think about the connection between the quadratic formula and 4.16.

We want a factorization theorem for polynomials with real coefficients. First we need to characterize the polynomials of degree 2 with real coefficients that can be written as the product of two polynomials of degree 1 with real coefficients.

4.16 Factorization of a quadratic polynomial

Suppose $b, c \in \mathbf{R}$. Then there is a polynomial factorization of the form

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

with $\lambda_1, \lambda_2 \in \mathbf{R}$ if and only if $b^2 \geq 4c$.

Proof Notice that

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right).$$

The equation above is the basis of the technique called ***completing the square****.*

First suppose $b^2 < 4c$. Then clearly the right side of the equation above is positive for every $x \in \mathbf{R}$. Hence the polynomial $x^2 + bx + c$ has no real zeros and thus cannot be factored in the form $(x-\lambda_1)(x-\lambda_2)$ with $\lambda_1, \lambda_2 \in \mathbf{R}$.

Conversely, now suppose $b^2 \geq 4c$. Then there is a real number d such that $d^2 = \frac{b^2}{4} - c$. From the displayed equation above, we have

$$\begin{aligned} x^2 + bx + c &= \left(x + \frac{b}{2}\right)^2 - d^2 \\ &= \left(x + \frac{b}{2} + d\right)\left(x + \frac{b}{2} - d\right), \end{aligned}$$

which gives the desired factorization. ■

The next result gives a factorization of a polynomial over $\mathbf{R}$. The idea of the proof is to use the factorization 4.14 of p as a polynomial with complex coefficients. Complex but nonreal zeros of p come in pairs; see 4.15. Thus if the factorization of p as an element of $\mathcal{P}(\mathbf{C})$ includes terms of the form $(x - \lambda)$ with λ a nonreal complex number, then $(x - \bar{\lambda})$ is also a term in the factorization. Multiplying together these two terms, we get

$$\left(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\right),$$

which is a quadratic term of the required form.

The idea sketched in the paragraph above almost provides a proof of the existence of our desired factorization. However, we need to be careful about one point. Suppose λ is a nonreal complex number and $(x - \lambda)$ is a term in the factorization of p as an element of $\mathcal{P}(\mathbf{C})$. We are guaranteed by 4.15 that $(x - \bar{\lambda})$ also appears as a term in the factorization, but 4.15 does not state that these two factors appear the same number of times, as needed to make the idea above work. However, the proof works around this point.

In the next result, either m or M may equal 0. The numbers $\lambda_1, \ldots, \lambda_m$ are precisely the real zeros of p, for these are the only real values of x for which the right side of the equation in the next result equals 0.

4.17 Factorization of a polynomial over $\mathbf{R}$

Suppose $p \in \mathcal{P}(\mathbf{R})$ is a nonconstant polynomial. Then p has a unique factorization (except for the order of the factors) of the form

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbf{R}$, with ${b_j}^2 < 4c_j$ for each j.

Proof Think of p as an element of $\mathcal{P}(\mathbf{C})$. If all the (complex) zeros of p are real, then we are done by 4.14. Thus suppose p has a zero $\lambda \in \mathbf{C}$ with $\lambda \notin \mathbf{R}$. By 4.15, $\bar{\lambda}$ is a zero of p. Thus we can write

$$\begin{aligned} p(x) &= (x - \lambda)(x - \bar{\lambda})q(x) \\ &= \big(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\big)q(x) \end{aligned}$$

for some polynomial $q \in \mathcal{P}(\mathbf{C})$ with degree two less than the degree of p. If we can prove that q has real coefficients, then by using induction on the degree of p, we can conclude that $(x - \lambda)$ appears in the factorization of p exactly as many times as $(x - \bar{\lambda})$.

To prove that q has real coefficients, we solve the equation above for q, getting

$$q(x) = \frac{p(x)}{x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2}$$

for all $x \in \mathbf{R}$. The equation above implies that $q(x) \in \mathbf{R}$ for all $x \in \mathbf{R}$. Writing

$$q(x) = a_0 + a_1 x + \cdots + a_{n-2}x^{n-2},$$

where $n = \deg p$ and $a_0, \ldots, a_{n-2} \in \mathbf{C}$, we thus have

$$0 = \operatorname{Im} q(x) = (\operatorname{Im} a_0) + (\operatorname{Im} a_1)x + \cdots + (\operatorname{Im} a_{n-2})x^{n-2}$$

for all $x \in \mathbf{R}$. This implies that $\operatorname{Im} a_0, \ldots, \operatorname{Im} a_{n-2}$ all equal 0 (by 4.7). Thus all the coefficients of q are real, as desired. Hence the desired factorization exists.

Now we turn to the question of uniqueness of our factorization. A factor of p of the form $x^2 + b_j x + c_j$ with ${b_j}^2 < 4c_j$ can be uniquely written as $(x - \lambda_j)(x - \overline{\lambda_j})$ with $\lambda_j \in \mathbf{C}$. A moment's thought shows that two different factorizations of p as an element of $\mathcal{P}(\mathbf{R})$ would lead to two different factorizations of p as an element of $\mathcal{P}(\mathbf{C})$, contradicting 4.14. ∎

EXERCISES 4

1 Verify all the assertions in 4.5 except the last one.

2 Suppose m is a positive integer. Is the set

$$\{0\} \cup \{p \in \mathcal{P}(\mathbf{F}) : \deg p = m\}$$

a subspace of $\mathcal{P}(\mathbf{F})$?

3 Is the set

$$\{0\} \cup \{p \in \mathcal{P}(\mathbf{F}) : \deg p \text{ is even}\}$$

a subspace of $\mathcal{P}(\mathbf{F})$?

4 Suppose m and n are positive integers with $m \leq n$, and suppose $\lambda_1, \dots, \lambda_m \in \mathbf{F}$. Prove that there exists a polynomial $p \in \mathcal{P}(\mathbf{F})$ with $\deg p = n$ such that $0 = p(\lambda_1) = \cdots = p(\lambda_m)$ and such that p has no other zeros.

5 Suppose m is a nonnegative integer, $z_1, \dots, z_{m+1}$ are distinct elements of $\mathbf{F}$, and $w_1, \dots, w_{m+1} \in \mathbf{F}$. Prove that there exists a unique polynomial $p \in \mathcal{P}_m(\mathbf{F})$ such that

$$p(z_j) = w_j$$

for $j = 1, \dots, m+1$.
[*This result can be proved without using linear algebra. However, try to find the clearer, shorter proof that uses some linear algebra.*]

6 Suppose $p \in \mathcal{P}(\mathbf{C})$ has degree m. Prove that p has m distinct zeros if and only if p and its derivative p' have no zeros in common.

7 Prove that every polynomial of odd degree with real coefficients has a real zero.

8 Define $T : \mathcal{P}(\mathbf{R}) \to \mathbf{R}^{\mathbf{R}}$ by

$$Tp = \begin{cases} \dfrac{p - p(3)}{x - 3} & \text{if } x \neq 3, \\[2ex] p'(3) & \text{if } x = 3. \end{cases}$$

Show that $Tp \in \mathcal{P}(\mathbf{R})$ for every polynomial $p \in \mathcal{P}(\mathbf{R})$ and that T is a linear map.

9 Suppose $p \in \mathcal{P}(\mathbf{C})$. Define $q\colon \mathbf{C} \to \mathbf{C}$ by

$$q(z) = p(z)\overline{p(\bar{z})}.$$

Prove that q is a polynomial with real coefficients.

10 Suppose m is a nonnegative integer and $p \in \mathcal{P}_m(\mathbf{C})$ is such that there exist distinct real numbers $x_0, x_1, \ldots, x_m$ such that $p(x_j) \in \mathbf{R}$ for $j = 0, 1, \ldots, m$. Prove that all the coefficients of p are real.

11 Suppose $p \in \mathcal{P}(\mathbf{F})$ with $p \neq 0$. Let $U = \{pq : q \in \mathcal{P}(\mathbf{F})\}$.

(a) Show that $\dim \mathcal{P}(\mathbf{F})/U = \deg p$.

(b) Find a basis of $\dim \mathcal{P}(\mathbf{F})/U$.

CHAPTER

5

Statue of Italian mathematician Leonardo of Pisa (1170–1250, approximate dates), also known as Fibonacci. Exercise 16 in Section 5.C shows how linear algebra can be used to find an explicit formula for the Fibonacci sequence.

Eigenvalues, Eigenvectors, and Invariant Subspaces

Linear maps from one vector space to another vector space were the objects of study in Chapter 3. Now we begin our investigation of linear maps from a finite-dimensional vector space to itself. Their study constitutes the most important part of linear algebra.

Our standing assumptions are as follows:

5.1 Notation $\mathbf{F}$, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a vector space over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- invariant subspaces
- eigenvalues, eigenvectors, and eigenspaces
- each operator on a finite-dimensional complex vector space has an eigenvalue and an upper-triangular matrix with respect to some basis

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_5

5.A *Invariant Subspaces*

In this chapter we develop the tools that will help us understand the structure of operators. Recall that an operator is a linear map from a vector space to itself. Recall also that we denote the set of operators on V by $\mathcal{L}(V)$; in other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

Let's see how we might better understand what an operator looks like. Suppose $T \in \mathcal{L}(V)$. If we have a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a proper subspace of V, then to understand the behavior of T, we need only understand the behavior of each $T|_{U_j}$; here $T|_{U_j}$ denotes the restriction of T to the smaller domain U_j. Dealing with $T|_{U_j}$ should be easier than dealing with T because U_j is a smaller vector space than V.

However, if we intend to apply tools useful in the study of operators (such as taking powers), then we have a problem: $T|_{U_j}$ may not map U_j into itself; in other words, $T|_{U_j}$ may not be an operator on U_j. Thus we are led to consider only decompositions of V of the form above where T maps each U_j into itself.

The notion of a subspace that gets mapped into itself is sufficiently important to deserve a name.

5.2 Definition ***invariant subspace***

Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called ***invariant*** under T if $u \in U$ implies $Tu \in U$.

In other words, U is invariant under T if $T|_U$ is an operator on U.

5.3 Example Suppose $T \in \mathcal{L}(V)$. Show that each of the following subspaces of V is invariant under T:

(a) $\{0\}$;

(b) V;

(c) null T;

(d) range T.

The most famous unsolved problem in functional analysis is called the ***invariant subspace problem****. It deals with invariant subspaces of operators on infinite-dimensional vector spaces.*

Solution

(a) If $u \in \{0\}$, then $u = 0$ and hence $Tu = 0 \in \{0\}$. Thus $\{0\}$ is invariant under T.

(b) If $u \in V$, then $Tu \in V$. Thus V is invariant under T.

(c) If $u \in \operatorname{null} T$, then $Tu = 0$, and hence $Tu \in \operatorname{null} T$. Thus $\operatorname{null} T$ is invariant under T.

(d) If $u \in \operatorname{range} T$, then $Tu \in \operatorname{range} T$. Thus $\operatorname{range} T$ is invariant under T.

Must an operator $T \in \mathcal{L}(V)$ have any invariant subspaces other than $\{0\}$ and V? Later we will see that this question has an affirmative answer if V is finite-dimensional and $\dim V > 1$ (for $\mathbf{F} = \mathbf{C}$) or $\dim V > 2$ (for $\mathbf{F} = \mathbf{R}$); see 5.21 and 9.8.

Although $\operatorname{null} T$ and $\operatorname{range} T$ are invariant under T, they do not necessarily provide easy answers to the question about the existence of invariant subspaces other than $\{0\}$ and V, because $\operatorname{null} T$ may equal $\{0\}$ and $\operatorname{range} T$ may equal V (this happens when T is invertible).

5.4 **Example** Suppose that $T \in \mathcal{L}\big(\mathcal{P}(\mathbf{R})\big)$ is defined by $Tp = p'$. Then $\mathcal{P}_4(\mathbf{R})$, which is a subspace of $\mathcal{P}(\mathbf{R})$, is invariant under T because if $p \in \mathcal{P}(\mathbf{R})$ has degree at most 4, then p' also has degree at most 4.

Eigenvalues and Eigenvectors

We will return later to a deeper study of invariant subspaces. Now we turn to an investigation of the simplest possible nontrivial invariant subspaces—invariant subspaces with dimension 1.

Take any $v \in V$ with $v \neq 0$ and let U equal the set of all scalar multiples of v:

$$U = \{\lambda v : \lambda \in \mathbf{F}\} = \operatorname{span}(v).$$

Then U is a 1-dimensional subspace of V (and every 1-dimensional subspace of V is of this form for an appropriate choice of v). If U is invariant under an operator $T \in \mathcal{L}(V)$, then $Tv \in U$, and hence there is a scalar $\lambda \in \mathbf{F}$ such that

$$Tv = \lambda v.$$

Conversely, if $Tv = \lambda v$ for some $\lambda \in \mathbf{F}$, then $\operatorname{span}(v)$ is a 1-dimensional subspace of V invariant under T.

The equation

$$Tv = \lambda v,$$

which we have just seen is intimately connected with 1-dimensional invariant subspaces, is important enough that the vectors v and scalars λ satisfying it are given special names.

5.5 Definition ***eigenvalue***

Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an ***eigenvalue*** of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

*The word **eigenvalue** is half-German, half-English. The German adjective **eigen** means "own" in the sense of characterizing an intrinsic property. Some mathematicians use the term **characteristic value** instead of eigenvalue.*

The comments above show that T has a 1-dimensional invariant subspace if and only if T has an eigenvalue.

In the definition above, we require that $v \neq 0$ because every scalar $\lambda \in \mathbf{F}$ satisfies $T0 = \lambda 0$.

5.6 Equivalent conditions to be an eigenvalue

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in F$. Then the following are equivalent:

(a) λ is an eigenvalue of T;

(b) $T - \lambda I$ is not injective;

(c) $T - \lambda I$ is not surjective;

(d) $T - \lambda I$ is not invertible.

Recall that $I \in \mathcal{L}(V)$ is the identity operator defined by $Iv = v$ for all $v \in V$.

Proof Conditions (a) and (b) are equivalent because the equation $Tv = \lambda v$ is equivalent to the equation $(T - \lambda I)v = 0$. Conditions (b), (c), and (d) are equivalent by 3.69. ■

5.7 Definition ***eigenvector***

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$ is an eigenvalue of T. A vector $v \in V$ is called an ***eigenvector*** of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

Because $Tv = \lambda v$ if and only if $(T - \lambda I)v = 0$, a vector $v \in V$ with $v \neq 0$ is an eigenvector of T corresponding to λ if and only if $v \in \operatorname{null}(T - \lambda I)$.

5.8 **Example** Suppose $T \in \mathcal{L}(\mathbf{F}^2)$ is defined by

$$T(w, z) = (-z, w).$$

(a) Find the eigenvalues and eigenvectors of T if $\mathbf{F} = \mathbf{R}$.

(b) Find the eigenvalues and eigenvectors of T if $\mathbf{F} = \mathbf{C}$.

Solution

(a) If $\mathbf{F} = \mathbf{R}$, then T is a counterclockwise rotation by 90° about the origin in $\mathbf{R}^2$. An operator has an eigenvalue if and only if there exists a nonzero vector in its domain that gets sent by the operator to a scalar multiple of itself. A 90° counterclockwise rotation of a nonzero vector in $\mathbf{R}^2$ obviously never equals a scalar multiple of itself. Conclusion: if $\mathbf{F} = \mathbf{R}$, then T has no eigenvalues (and thus has no eigenvectors).

(b) To find eigenvalues of T, we must find the scalars λ such that

$$T(w, z) = \lambda(w, z)$$

has some solution other than $w = z = 0$. The equation above is equivalent to the simultaneous equations

5.9 $$-z = \lambda w, \quad w = \lambda z.$$

Substituting the value for w given by the second equation into the first equation gives

$$-z = \lambda^2 z.$$

Now z cannot equal 0 [otherwise 5.9 implies that $w = 0$; we are looking for solutions to 5.9 where (w, z) is not the 0 vector], so the equation above leads to the equation

$$-1 = \lambda^2.$$

The solutions to this equation are $\lambda = i$ and $\lambda = -i$. You should be able to verify easily that i and $-i$ are eigenvalues of T. Indeed, the eigenvectors corresponding to the eigenvalue i are the vectors of the form $(w, -wi)$, with $w \in \mathbf{C}$ and $w \neq 0$, and the eigenvectors corresponding to the eigenvalue $-i$ are the vectors of the form (w, wi), with $w \in \mathbf{C}$ and $w \neq 0$.

Now we show that eigenvectors corresponding to distinct eigenvalues are linearly independent.

5.10 Linearly independent eigenvectors

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.

Proof Suppose $v_1, \dots, v_m$ is linearly dependent. Let k be the smallest positive integer such that

5.11 $$v_k \in \operatorname{span}(v_1, \dots, v_{k-1});$$

the existence of k with this property follows from the Linear Dependence Lemma (2.21). Thus there exist $a_1, \dots, a_{k-1} \in \mathbf{F}$ such that

5.12 $$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}.$$

Apply T to both sides of this equation, getting

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1}.$$

Multiply both sides of 5.12 by λ_k and then subtract the equation above, getting

$$0 = a_1(\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1}) v_{k-1}.$$

Because we chose k to be the smallest positive integer satisfying 5.11, $v_1, \dots, v_{k-1}$ is linearly independent. Thus the equation above implies that all the a's are 0 (recall that λ_k is not equal to any of $\lambda_1, \dots, \lambda_{k-1}$). However, this means that v_k equals 0 (see 5.12), contradicting our hypothesis that v_k is an eigenvector. Therefore our assumption that $v_1, \dots, v_m$ is linearly dependent was false. ■

The corollary below states that an operator cannot have more distinct eigenvalues than the dimension of the vector space on which it acts.

5.13 Number of eigenvalues

Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Proof Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T. Let $v_1, \dots, v_m$ be corresponding eigenvectors. Then 5.10 implies that the list $v_1, \dots, v_m$ is linearly independent. Thus $m \leq \dim V$ (see 2.23), as desired. ■

Restriction and Quotient Operators

If $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T, then U determines two other operators $T|_U \in \mathcal{L}(U)$ and $T/U \in \mathcal{L}(V/U)$ in a natural way, as defined below.

5.14 Definition $\quad T|_U$ ***and*** T/U

Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T.

- The ***restriction operator*** $T|_U \in \mathcal{L}(U)$ is defined by
$$T|_U(u) = Tu$$
for $u \in U$.
- The ***quotient operator*** $T/U \in \mathcal{L}(V/U)$ is defined by
$$(T/U)(v + U) = Tv + U$$
for $v \in V$.

For both the operators defined above, it is worthwhile to pay attention to their domains and to spend a moment thinking about why they are well defined as operators on their domains. First consider the restriction operator $T|_U \in \mathcal{L}(U)$, which is T with its domain restricted to U, thought of as mapping into U instead of into V. The condition that U is invariant under T is what allows us to think of $T|_U$ as an operator on U, meaning a linear map into the same space as the domain, rather than as simply a linear map from one vector space to another vector space.

To show that the definition above of the quotient operator makes sense, we need to verify that if $v + U = w + U$, then $Tv + U = Tw + U$. Hence suppose $v + U = w + U$. Thus $v - w \in U$ (see 3.85). Because U is invariant under T, we also have $T(v - w) \in U$, which implies that $Tv - Tw \in U$, which implies that $Tv + U = Tw + U$, as desired.

Suppose T is an operator on a finite-dimensional vector space V and U is a subspace of V invariant under T, with $U \neq \{0\}$ and $U \neq V$. In some sense, we can learn about T by studying the operators $T|_U$ and T/U, each of which is an operator on a vector space with smaller dimension than V. For example, proof 2 of 5.27 makes nice use of T/U.

However, sometimes $T|_U$ and T/U do not provide enough information about T. In the next example, both $T|_U$ and T/U are 0 even though T is not the 0 operator.

5.15 **Example** Define an operator $T \in \mathcal{L}(\mathbf{F}^2)$ by $T(x, y) = (y, 0)$. Let $U = \{(x, 0) : x \in \mathbf{F}\}$. Show that

(a) U is invariant under T and $T|_U$ is the 0 operator on U;

(b) there does not exist a subspace W of $\mathbf{F}^2$ that is invariant under T and such that $\mathbf{F}^2 = U \oplus W$;

(c) T/U is the 0 operator on $\mathbf{F}^2/U$.

Solution

(a) For $(x, 0) \in U$, we have $T(x, 0) = (0, 0) \in U$. Thus U is invariant under T and $T|_U$ is the 0 operator on U.

(b) Suppose W is a subspace of V such that $\mathbf{F}^2 = U \oplus W$. Because $\dim \mathbf{F}^2 = 2$ and $\dim U = 1$, we have $\dim W = 1$. If W were invariant under T, then each nonzero vector in W would be an eigenvector of T. However, it is easy to see that 0 is the only eigenvalue of T and that all eigenvectors of T are in U. Thus W is not invariant under T.

(c) For $(x, y) \in \mathbf{F}^2$, we have

$$\begin{aligned}(T/U)\big((x, y) + U\big) &= T(x, y) + U \\ &= (y, 0) + U \\ &= 0 + U,\end{aligned}$$

where the last equality holds because $(y, 0) \in U$. The equation above shows that T/U is the 0 operator.

EXERCISES 5.A

1 Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V.

(a) Prove that if $U \subset \operatorname{null} T$, then U is invariant under T.

(b) Prove that if $\operatorname{range} T \subset U$, then U is invariant under T.

2 Suppose $S, T \in \mathcal{L}(V)$ are such that $ST = TS$. Prove that $\operatorname{null} S$ is invariant under T.

3 Suppose $S, T \in \mathcal{L}(V)$ are such that $ST = TS$. Prove that range S is invariant under T.

4 Suppose that $T \in \mathcal{L}(V)$ and $U_1, \dots, U_m$ are subspaces of V invariant under T. Prove that $U_1 + \cdots + U_m$ is invariant under T.

5 Suppose $T \in \mathcal{L}(V)$. Prove that the intersection of every collection of subspaces of V invariant under T is invariant under T.

6 Prove or give a counterexample: if V is finite-dimensional and U is a subspace of V that is invariant under every operator on V, then $U = \{0\}$ or $U = V$.

7 Suppose $T \in \mathcal{L}(\mathbf{R}^2)$ is defined by $T(x, y) = (-3y, x)$. Find the eigenvalues of T.

8 Define $T \in \mathcal{L}(\mathbf{F}^2)$ by

$$T(w, z) = (z, w).$$

Find all eigenvalues and eigenvectors of T.

9 Define $T \in \mathcal{L}(\mathbf{F}^3)$ by

$$T(z_1, z_2, z_3) = (2z_2, 0, 5z_3).$$

Find all eigenvalues and eigenvectors of T.

10 Define $T \in \mathcal{L}(\mathbf{F}^n)$ by

$$T(x_1, x_2, x_3, \dots, x_n) = (x_1, 2x_2, 3x_3, \dots, nx_n).$$

(a) Find all eigenvalues and eigenvectors of T.

(b) Find all invariant subspaces of T.

11 Define $T : \mathcal{P}(\mathbf{R}) \to \mathcal{P}(\mathbf{R})$ by $Tp = p'$. Find all eigenvalues and eigenvectors of T.

12 Define $T \in \mathcal{L}\big(\mathcal{P}_4(\mathbf{R})\big)$ by

$$(Tp)(x) = xp'(x)$$

for all $x \in \mathbf{R}$. Find all eigenvalues and eigenvectors of T.

13 Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{F}$. Prove that there exists $\alpha \in \mathbf{F}$ such that $|\alpha - \lambda| < \frac{1}{1000}$ and $T - \alpha I$ is invertible.

14 Suppose $V = U \oplus W$, where U and W are nonzero subspaces of V. Define $P \in \mathcal{L}(V)$ by $P(u + w) = u$ for $u \in U$ and $w \in W$. Find all eigenvalues and eigenvectors of P.

15 Suppose $T \in \mathcal{L}(V)$. Suppose $S \in \mathcal{L}(V)$ is invertible.

(a) Prove that T and $S^{-1}TS$ have the same eigenvalues.

(b) What is the relationship between the eigenvectors of T and the eigenvectors of $S^{-1}TS$?

16 Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and the matrix of T with respect to some basis of V contains only real entries. Show that if λ is an eigenvalue of T, then so is $\bar{\lambda}$.

17 Give an example of an operator $T \in \mathcal{L}(\mathbf{R}^4)$ such that T has no (real) eigenvalues.

18 Show that the operator $T \in \mathcal{L}(\mathbf{C}^\infty)$ defined by

$$T(z_1, z_2, \dots) = (0, z_1, z_2, \dots)$$

has no eigenvalues.

19 Suppose n is a positive integer and $T \in \mathcal{L}(\mathbf{F}^n)$ is defined by

$$T(x_1, \dots, x_n) = (x_1 + \cdots + x_n, \dots, x_1 + \cdots + x_n);$$

in other words, T is the operator whose matrix (with respect to the standard basis) consists of all 1's. Find all eigenvalues and eigenvectors of T.

20 Find all eigenvalues and eigenvectors of the backward shift operator $T \in \mathcal{L}(\mathbf{F}^\infty)$ defined by

$$T(z_1, z_2, z_3, \dots) = (z_2, z_3, \dots).$$

21 Suppose $T \in \mathcal{L}(V)$ is invertible.

(a) Suppose $\lambda \in \mathbf{F}$ with $\lambda \neq 0$. Prove that λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}.

(b) Prove that T and T^{-1} have the same eigenvectors.

22 Suppose $T \in \mathcal{L}(V)$ and there exist nonzero vectors v and w in V such that

$$Tv = 3w \quad \text{and} \quad Tw = 3v.$$

Prove that 3 or -3 is an eigenvalue of T.

23 Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST and TS have the same eigenvalues.

24 Suppose A is an n-by-n matrix with entries in $\mathbf{F}$. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by $Tx = Ax$, where elements of $\mathbf{F}^n$ are thought of as n-by-1 column vectors.

(a) Suppose the sum of the entries in each row of A equals 1. Prove that 1 is an eigenvalue of T.

(b) Suppose the sum of the entries in each column of A equals 1. Prove that 1 is an eigenvalue of T.

25 Suppose $T \in \mathcal{L}(V)$ and u, v are eigenvectors of T such that $u + v$ is also an eigenvector of T. Prove that u and v are eigenvectors of T corresponding to the same eigenvalue.

26 Suppose $T \in \mathcal{L}(V)$ is such that every nonzero vector in V is an eigenvector of T. Prove that T is a scalar multiple of the identity operator.

27 Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$ is such that every subspace of V with dimension $\dim V - 1$ is invariant under T. Prove that T is a scalar multiple of the identity operator.

28 Suppose V is finite-dimensional with $\dim V \geq 3$ and $T \in \mathcal{L}(V)$ is such that every 2-dimensional subspace of V is invariant under T. Prove that T is a scalar multiple of the identity operator.

29 Suppose $T \in \mathcal{L}(V)$ and $\dim \operatorname{range} T = k$. Prove that T has at most $k + 1$ distinct eigenvalues.

30 Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ and -4, 5, and $\sqrt{7}$ are eigenvalues of T. Prove that there exists $x \in \mathbf{R}^3$ such that $Tx - 9x = (-4, 5, \sqrt{7})$.

31 Suppose V is finite-dimensional and $v_1, \dots, v_m$ is a list of vectors in V. Prove that $v_1, \dots, v_m$ is linearly independent if and only if there exists $T \in \mathcal{L}(V)$ such that $v_1, \dots, v_m$ are eigenvectors of T corresponding to distinct eigenvalues.

32 Suppose $\lambda_1, \ldots, \lambda_n$ is a list of distinct real numbers. Prove that the list $e^{\lambda_1 x}, \ldots, e^{\lambda_n x}$ is linearly independent in the vector space of real-valued functions on $\mathbf{R}$.

Hint: Let $V = \operatorname{span}\bigl(e^{\lambda_1 x}, \ldots, e^{\lambda_n x}\bigr)$, and define an operator $T \in \mathcal{L}(V)$ by $Tf = f'$. Find eigenvalues and eigenvectors of T.

33 Suppose $T \in \mathcal{L}(V)$. Prove that $T/(\operatorname{range} T) = 0$.

34 Suppose $T \in \mathcal{L}(V)$. Prove that $T/(\operatorname{null} T)$ is injective if and only if $(\operatorname{null} T) \cap (\operatorname{range} T) = \{0\}$.

35 Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is invariant under T. Prove that each eigenvalue of T/U is an eigenvalue of T.
[*The exercise below asks you to verify that the hypothesis that V is finite-dimensional is needed for the exercise above.*]

36 Give an example of a vector space V, an operator $T \in \mathcal{L}(V)$, and a subspace U of V that is invariant under T such that T/U has an eigenvalue that is not an eigenvalue of T.

5.B *Eigenvectors and Upper-Triangular Matrices*

Polynomials Applied to Operators

The main reason that a richer theory exists for operators (which map a vector space into itself) than for more general linear maps is that operators can be raised to powers. We begin this section by defining that notion and the key concept of applying a polynomial to an operator.

If $T \in \mathcal{L}(V)$, then TT makes sense and is also in $\mathcal{L}(V)$. We usually write T^2 instead of TT. More generally, we have the following definition.

5.16 Definition T^m

Suppose $T \in \mathcal{L}(V)$ and m is a positive integer.

- T^m is defined by
$$T^m = \underbrace{T \cdots T}_{m \text{ times}}.$$
- T^0 is defined to be the identity operator I on V.
- If T is invertible with inverse T^{-1}, then T^{-m} is defined by
$$T^{-m} = (T^{-1})^m.$$

You should verify that if T is an operator, then

$$T^m T^n = T^{m+n} \quad \text{and} \quad (T^m)^n = T^{mn},$$

where m and n are allowed to be arbitrary integers if T is invertible and nonnegative integers if T is not invertible.

5.17 Definition $p(T)$

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$ is a polynomial given by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for $z \in \mathbf{F}$. Then $p(T)$ is the operator defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

This is a new use of the symbol p because we are applying it to operators, not just elements of $\mathbf{F}$.

5.18 Example Suppose $D \in \mathcal{L}(\mathcal{P}(\mathbf{R}))$ is the differentiation operator defined by $Dq = q'$ and p is the polynomial defined by $p(x) = 7 - 3x + 5x^2$. Then $p(D) = 7I - 3D + 5D^2$; thus

$$\big(p(D)\big)q = 7q - 3q' + 5q''$$

for every $q \in \mathcal{P}(\mathbf{R})$.

If we fix an operator $T \in \mathcal{L}(V)$, then the function from $\mathcal{P}(\mathbf{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear, as you should verify.

5.19 Definition ***product of polynomials***

If $p, q \in \mathcal{P}(\mathbf{F})$, then $pq \in \mathcal{P}(\mathbf{F})$ is the polynomial defined by

$$(pq)(z) = p(z)q(z)$$

for $z \in \mathbf{F}$.

Any two polynomials of an operator commute, as shown below.

5.20 Multiplicative properties

Suppose $p, q \in \mathcal{P}(\mathbf{F})$ and $T \in \mathcal{L}(V)$. Then

(a) $(pq)(T) = p(T)q(T)$;

(b) $p(T)q(T) = q(T)p(T)$.

Part (a) holds because when expanding a product of polynomials using the distributive property, it does not matter whether the symbol is z or T.

Proof

(a) Suppose $p(z) = \sum_{j=0}^{m} a_j z^j$ and $q(z) = \sum_{k=0}^{n} b_k z^k$ for $z \in \mathbf{F}$. Then

$$(pq)(z) = \sum_{j=0}^{m} \sum_{k=0}^{n} a_j b_k z^{j+k}.$$

Thus

$$\begin{aligned}(pq)(T) &= \sum_{j=0}^{m} \sum_{k=0}^{n} a_j b_k T^{j+k} \\ &= \Big(\sum_{j=0}^{m} a_j T^j\Big)\Big(\sum_{k=0}^{n} b_k T^k\Big) \\ &= p(T)q(T).\end{aligned}$$

(b) Part (a) implies $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$. ■

Existence of Eigenvalues

Now we come to one of the central results about operators on complex vector spaces.

5.21 Operators on complex vector spaces have an eigenvalue

Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

Proof Suppose V is a complex vector space with dimension $n > 0$ and $T \in \mathcal{L}(V)$. Choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2v, \ldots, T^nv$$

is not linearly independent, because V has dimension n and we have $n + 1$ vectors. Thus there exist complex numbers $a_0, \ldots, a_n$, not all 0, such that

$$0 = a_0v + a_1Tv + \cdots + a_nT^nv.$$

Note that $a_1, \ldots, a_n$ cannot all be 0, because otherwise the equation above would become $0 = a_0v$, which would force a_0 also to be 0.

Make the a's the coefficients of a polynomial, which by the Fundamental Theorem of Algebra (4.14) has a factorization

$$a_0 + a_1z + \cdots + a_nz^n = c(z - \lambda_1)\cdots(z - \lambda_m),$$

where c is a nonzero complex number, each λ_j is in $\mathbf{C}$, and the equation holds for all $z \in \mathbf{C}$ (here m is not necessarily equal to n, because a_n may equal 0). We then have

$$\begin{aligned} 0 &= a_0v + a_1Tv + \cdots + a_nT^nv \\ &= (a_0I + a_1T + \cdots + a_nT^n)v \\ &= c(T - \lambda_1I)\cdots(T - \lambda_mI)v. \end{aligned}$$

Thus $T - \lambda_jI$ is not injective for at least one j. In other words, T has an eigenvalue. ■

The proof above depends on the Fundamental Theorem of Algebra, which is typical of proofs of this result. See Exercises 16 and 17 for possible ways to rewrite the proof above using the idea of the proof in a slightly different form.

Upper-Triangular Matrices

In Chapter 3 we discussed the matrix of a linear map from one vector space to another vector space. That matrix depended on a choice of a basis of each of the two vector spaces. Now that we are studying operators, which map a vector space to itself, the emphasis is on using only one basis.

5.22 Definition ***matrix of an operator,*** $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V. The ***matrix of*** T with respect to this basis is the n-by-n matrix

$$\mathcal{M}(T) = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}$$

whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}v_1 + \dots + A_{n,k}v_n.$$

If the basis is not clear from the context, then the notation $\mathcal{M}\big(T, (v_1, \dots, v_n)\big)$ is used.

Note that the matrices of operators are square arrays, rather than the more general rectangular arrays that we considered earlier for linear maps.

The k^{th} column of the matrix $\mathcal{M}(T)$ is formed from the coefficients used to write Tv_k as a linear combination of $v_1, \dots, v_n$.

If T is an operator on $\mathbf{F}^n$ and no basis is specified, assume that the basis in question is the standard one (where the j^{th} basis vector is 1 in the j^{th} slot and 0 in all the other slots). You can then think of the j^{th} column of $\mathcal{M}(T)$ as T applied to the j^{th} basis vector.

5.23 Example Define $T \in \mathcal{L}(\mathbf{F}^3)$ by $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. Then

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

A central goal of linear algebra is to show that given an operator $T \in \mathcal{L}(V)$, there exists a basis of V with respect to which T has a reasonably simple matrix. To make this vague formulation a bit more precise, we might try to choose a basis of V such that $\mathcal{M}(T)$ has many 0's.

If V is a finite-dimensional complex vector space, then we already know enough to show that there is a basis of V with respect to which the matrix of T has 0's everywhere in the first column, except possibly the first entry. In other words, there is a basis of V with respect to which the matrix of T looks like

$$\begin{pmatrix} \lambda & & & \\ 0 & & * & \\ \vdots & & & \\ 0 & & & \end{pmatrix};$$

here the $*$ denotes the entries in all the columns other than the first column. To prove this, let λ be an eigenvalue of T (one exists by 5.21) and let v be a corresponding eigenvector. Extend v to a basis of V. Then the matrix of T with respect to this basis has the form above.

Soon we will see that we can choose a basis of V with respect to which the matrix of T has even more 0's.

5.24 Definition ***diagonal of a matrix***

The ***diagonal*** of a square matrix consists of the entries along the line from the upper left corner to the bottom right corner.

For example, the diagonal of the matrix in 5.23 consists of the entries $2, 5, 8$.

5.25 Definition ***upper-triangular matrix***

A matrix is called ***upper triangular*** if all the entries below the diagonal equal 0.

For example, the matrix in 5.23 is upper triangular.

Typically we represent an upper-triangular matrix in the form

$$\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix};$$

the 0 in the matrix above indicates that all entries below the diagonal in this n-by-n matrix equal 0. Upper-triangular matrices can be considered reasonably simple—for n large, almost half its entries in an n-by-n upper-triangular matrix are 0.

We often use $*$ to denote matrix entries that we do not know about or that are irrelevant to the questions being discussed.

The following proposition demonstrates a useful connection between upper-triangular matrices and invariant subspaces.

5.26 Conditions for upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following are equivalent:

(a) the matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular;

(b) $Tv_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each $j = 1, \ldots, n$;

(c) $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $j = 1, \ldots, n$.

Proof The equivalence of (a) and (b) follows easily from the definitions and a moment's thought. Obviously (c) implies (b). Hence to complete the proof, we need only prove that (b) implies (c).

Thus suppose (b) holds. Fix $j \in \{1, \ldots, n\}$. From (b), we know that

$$\begin{aligned} Tv_1 &\in \operatorname{span}(v_1) \subset \operatorname{span}(v_1, \ldots, v_j); \\ Tv_2 &\in \operatorname{span}(v_1, v_2) \subset \operatorname{span}(v_1, \ldots, v_j); \\ &\vdots \\ Tv_j &\in \operatorname{span}(v_1, \ldots, v_j). \end{aligned}$$

Thus if v is a linear combination of $v_1, \ldots, v_j$, then

$$Tv \in \operatorname{span}(v_1, \ldots, v_j).$$

In other words, $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T, completing the proof. ■

The next result does not hold on real vector spaces, because the first vector in a basis with respect to which an operator has an upper-triangular matrix is an eigenvector of the operator. Thus if an operator on a real vector space has no eigenvalues [see 5.8(a) for an example], then there is no basis with respect to which the operator has an upper-triangular matrix.

Now we can prove that for each operator on a finite-dimensional complex vector space, there is a basis of the vector space with respect to which the matrix of the operator has only 0's below the diagonal. In Chapter 8 we will improve even this result.

Sometimes more insight comes from seeing more than one proof of a theorem. Thus two proofs are presented of the next result. Use whichever appeals more to you.

5.27 Over $\mathbf{C}$, every operator has an upper-triangular matrix

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

Proof 1 We will use induction on the dimension of V. Clearly the desired result holds if $\dim V = 1$.

Suppose now that $\dim V > 1$ and the desired result holds for all complex vector spaces whose dimension is less than the dimension of V. Let λ be any eigenvalue of T (5.21 guarantees that T has an eigenvalue). Let

$$U = \operatorname{range}(T - \lambda I).$$

Because $T - \lambda I$ is not surjective (see 3.69), $\dim U < \dim V$. Furthermore, U is invariant under T. To prove this, suppose $u \in U$. Then

$$Tu = (T - \lambda I)u + \lambda u.$$

Obviously $(T - \lambda I)u \in U$ (because U equals the range of $T - \lambda I$) and $\lambda u \in U$. Thus the equation above shows that $Tu \in U$. Hence U is invariant under T, as claimed.

Thus $T|_U$ is an operator on U. By our induction hypothesis, there is a basis $u_1, \dots, u_m$ of U with respect to which $T|_U$ has an upper-triangular matrix. Thus for each j we have (using 5.26)

5.28 $$Tu_j = (T|_U)(u_j) \in \operatorname{span}(u_1, \dots, u_j).$$

Extend $u_1, \dots, u_m$ to a basis $u_1, \dots, u_m, v_1, \dots, v_n$ of V. For each k, we have

$$Tv_k = (T - \lambda I)v_k + \lambda v_k.$$

The definition of U shows that $(T - \lambda I)v_k \in U = \operatorname{span}(u_1, \dots, u_m)$. Thus the equation above shows that

5.29 $$Tv_k \in \operatorname{span}(u_1, \dots, u_m, v_1, \dots, v_k).$$

From 5.28 and 5.29, we conclude (using 5.26) that T has an upper-triangular matrix with respect to the basis $u_1, \dots, u_m, v_1, \dots, v_n$ of V, as desired. ■

Proof 2 We will use induction on the dimension of V. Clearly the desired result holds if $\dim V = 1$.

Suppose now that $\dim V = n > 1$ and the desired result holds for all complex vector spaces whose dimension is $n - 1$. Let v_1 be any eigenvector of T (5.21 guarantees that T has an eigenvector). Let $U = \operatorname{span}(v_1)$. Then U is an invariant subspace of T and $\dim U = 1$.

Because $\dim V/U = n - 1$ (see 3.89), we can apply our induction hypothesis to $T/U \in \mathcal{L}(V/U)$. Thus there is a basis $v_2 + U, \dots, v_n + U$ of V/U such that T/U has an upper-triangular matrix with respect to this basis. Hence by 5.26,

$$(T/U)(v_j + U) \in \operatorname{span}(v_2 + U, \dots, v_j + U)$$

for each $j = 2, \dots, n$. Unraveling the meaning of the inclusion above, we see that

$$Tv_j \in \operatorname{span}(v_1, \dots, v_j)$$

for each $j = 1, \dots, n$. Thus by 5.26, T has an upper-triangular matrix with respect to the basis $v_1, \dots, v_n$ of V, as desired (it is easy to verify that $v_1, \dots, v_n$ is a basis of V; see Exercise 13 in Section 3.E for a more general result). ■

How does one determine from looking at the matrix of an operator whether the operator is invertible? If we are fortunate enough to have a basis with respect to which the matrix of the operator is upper triangular, then this problem becomes easy, as the following proposition shows.

5.30 Determination of invertibility from upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.

Proof Suppose $v_1, \dots, v_n$ is a basis of V with respect to which T has an upper-triangular matrix

5.31
$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

We need to prove that T is invertible if and only if all the λ_j's are nonzero.

First suppose the diagonal entries $\lambda_1, \dots, \lambda_n$ are all nonzero. The upper-triangular matrix in 5.31 implies that $Tv_1 = \lambda_1 v_1$. Because $\lambda_1 \neq 0$, we have $T(v_1/\lambda_1) = v_1$; thus $v_1 \in \operatorname{range} T$.

Now

$$T(v_2/\lambda_2) = av_1 + v_2$$

for some $a \in \mathbf{F}$. The left side of the equation above and av_1 are both in $\operatorname{range} T$; thus $v_2 \in \operatorname{range} T$.

Similarly, we see that

$$T(v_3/\lambda_3) = bv_1 + cv_2 + v_3$$

for some $b, c \in \mathbf{F}$. The left side of the equation above and bv_1, cv_2 are all in $\operatorname{range} T$; thus $v_3 \in \operatorname{range} T$.

Continuing in this fashion, we conclude that $v_1, \dots, v_n \in \operatorname{range} T$. Because $v_1, \dots, v_n$ is a basis of V, this implies that $\operatorname{range} T = V$. In other words, T is surjective. Hence T is invertible (by 3.69), as desired.

To prove the other direction, now suppose that T is invertible. This implies that $\lambda_1 \neq 0$, because otherwise we would have $Tv_1 = 0$.

Let $1 < j \leq n$, and suppose $\lambda_j = 0$. Then 5.31 implies that T maps $\operatorname{span}(v_1, \dots, v_j)$ into $\operatorname{span}(v_1, \dots, v_{j-1})$. Because

$$\dim \operatorname{span}(v_1, \dots, v_j) = j \quad \text{and} \quad \dim \operatorname{span}(v_1, \dots, v_{j-1}) = j - 1,$$

this implies that T restricted to $\dim \operatorname{span}(v_1, \dots, v_j)$ is not injective (by 3.23). Thus there exists $v \in \operatorname{span}(v_1, \dots, v_j)$ such that $v \neq 0$ and $Tv = 0$. Thus T is not injective, which contradicts our hypothesis (for this direction) that T is invertible. This contradiction means that our assumption that $\lambda_j = 0$ must be false. Hence $\lambda_j \neq 0$, as desired. ■

As an example of the result above, we see that the operator in Example 5.23 is invertible.

Unfortunately no method exists for exactly computing the eigenvalues of an operator from its matrix. However, if we are fortunate enough to find a basis with respect to which the matrix of the operator is upper triangular, then the problem of computing the eigenvalues becomes trivial, as the following proposition shows.

Powerful numeric techniques exist for finding good approximations to the eigenvalues of an operator from its matrix.

5.32 Determination of eigenvalues from upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

Proof Suppose $v_1, \dots, v_n$ is a basis of V with respect to which T has an upper-triangular matrix

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Let $\lambda \in \mathbf{F}$. Then

$$\mathcal{M}(T - \lambda I) = \begin{pmatrix} \lambda_1 - \lambda & & & * \\ & \lambda_2 - \lambda & & \\ & & \ddots & \\ 0 & & & \lambda_n - \lambda \end{pmatrix}.$$

Hence $T - \lambda I$ is not invertible if and only if λ equals one of the numbers $\lambda_1, \dots, \lambda_n$ (by 5.30). Thus λ is an eigenvalue of T if and only if λ equals one of the numbers $\lambda_1, \dots, \lambda_n$. ■

5.33 **Example** Define $T \in \mathcal{L}(\mathbf{F}^3)$ by $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. What are the eigenvalues of T?

Solution The matrix of T with respect to the standard basis is

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

Thus $\mathcal{M}(T)$ is an upper-triangular matrix. Now 5.32 implies that the eigenvalues of T are 2, 5, and 8.

Once the eigenvalues of an operator on $\mathbf{F}^n$ are known, the eigenvectors can be found easily using Gaussian elimination.

EXERCISES 5.B

1 Suppose $T \in \mathcal{L}(V)$ and there exists a positive integer n such that $T^n = 0$.

(a) Prove that $I - T$ is invertible and that

$$(I - T)^{-1} = I + T + \cdots + T^{n-1}.$$

(b) Explain how you would guess the formula above.

2 Suppose $T \in \mathcal{L}(V)$ and $(T - 2I)(T - 3I)(T - 4I) = 0$. Suppose λ is an eigenvalue of T. Prove that $\lambda = 2$ or $\lambda = 3$ or $\lambda = 4$.

3 Suppose $T \in \mathcal{L}(V)$ and $T^2 = I$ and -1 is not an eigenvalue of T. Prove that $T = I$.

4 Suppose $P \in \mathcal{L}(V)$ and $P^2 = P$. Prove that $V = \operatorname{null} P \oplus \operatorname{range} P$.

v = Pv + (v - Pv)

5 Suppose $S, T \in \mathcal{L}(V)$ and S is invertible. Suppose $p \in \mathcal{P}(\mathbf{F})$ is a polynomial. Prove that

$$p(STS^{-1}) = Sp(T)S^{-1}.$$

6 Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T. Prove that U is invariant under $p(T)$ for every polynomial $p \in \mathcal{P}(\mathbf{F})$.

7 Suppose $T \in \mathcal{L}(V)$. Prove that 9 is an eigenvalue of T^2 if and only if 3 or -3 is an eigenvalue of T.

8 Give an example of $T \in \mathcal{L}(\mathbf{R}^2)$ such that $T^4 = -1$.

9 Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $v \in V$ with $v \neq 0$. Let p be a nonzero polynomial of smallest degree such that $p(T)v = 0$. Prove that every zero of p is an eigenvalue of T.

10 Suppose $T \in \mathcal{L}(V)$ and v is an eigenvector of T with eigenvalue λ. Suppose $p \in \mathcal{P}(\mathbf{F})$. Prove that $p(T)v = p(\lambda)v$.

11 Suppose $\mathbf{F} = \mathbf{C}$, $T \in \mathcal{L}(V)$, $p \in \mathcal{P}(\mathbf{C})$ is a polynomial, and $\alpha \in \mathbf{C}$. Prove that α is an eigenvalue of $p(T)$ if and only if $\alpha = p(\lambda)$ for some eigenvalue λ of T.

12 Show that the result in the previous exercise does not hold if $\mathbf{C}$ is replaced with $\mathbf{R}$.

(x, y) -> (-y, x) => T^2 = - I

13 Suppose W is a complex vector space and $T \in \mathcal{L}(W)$ has no eigenvalues. Prove that every subspace of W invariant under T is either $\{0\}$ or infinite-dimensional.

14 Give an example of an operator whose matrix with respect to some basis contains only 0's on the diagonal, but the operator is invertible.
[*The exercise above and the exercise below show that 5.30 fails without the hypothesis that an upper-triangular matrix is under consideration.*]

15 Give an example of an operator whose matrix with respect to some basis contains only nonzero numbers on the diagonal, but the operator is not invertible.

16 Rewrite the proof of 5.21 using the linear map that sends $p \in \mathcal{P}_n(\mathbf{C})$ to $\big(p(T)\big)v \in V$ (and use 3.23).

17 Rewrite the proof of 5.21 using the linear map that sends $p \in \mathcal{P}_{n^2}(\mathbf{C})$ to $p(T) \in \mathcal{L}(V)$ (and use 3.23).

18 Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Define a function $f : \mathbf{C} \to \mathbf{R}$ by

$$f(\lambda) = \dim \operatorname{range}(T - \lambda I).$$

Prove that f is not a continuous function.

19 Suppose V is finite-dimensional with $\dim V > 1$ and $T \in \mathcal{L}(V)$. Prove that

$$\{p(T) : p \in \mathcal{P}(\mathbf{F})\} \neq \mathcal{L}(V).$$

20 Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Prove that T has an invariant subspace of dimension k for each $k = 1, \ldots, \dim V$.

5.C Eigenspaces and Diagonal Matrices

5.34 Definition ***diagonal matrix***

A ***diagonal matrix*** is a square matrix that is 0 everywhere except possibly along the diagonal.

5.35 Example

$$\begin{pmatrix} 8 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

is a diagonal matrix.

Obviously every diagonal matrix is upper triangular. In general, a diagonal matrix has many more 0's than an upper-triangular matrix.

If an operator has a diagonal matrix with respect to some basis, then the entries along the diagonal are precisely the eigenvalues of the operator; this follows from 5.32 (or find an easier proof for diagonal matrices).

5.36 Definition ***eigenspace,*** $E(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The ***eigenspace*** of T corresponding to λ, denoted $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \operatorname{null}(T - \lambda I).$$

In other words, $E(\lambda, T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

For $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, the eigenspace $E(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V). The definitions imply that λ is an eigenvalue of T if and only if $E(\lambda, T) \neq \{0\}$.

5.37 Example Suppose the matrix of an operator $T \in \mathcal{L}(V)$ with respect to a basis v_1, v_2, v_3 of V is the matrix in Example 5.35 above. Then

$$E(8, T) = \operatorname{span}(v_1), \quad E(5, T) = \operatorname{span}(v_2, v_3).$$

If λ is an eigenvalue of an operator $T \in \mathcal{L}(V)$, then T restricted to $E(\lambda, T)$ is just the operator of multiplication by λ.

5.38 Sum of eigenspaces is a direct sum

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Furthermore,

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

Proof To show that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, suppose

$$u_1 + \cdots + u_m = 0,$$

where each u_j is in $E(\lambda, T)$. Because eigenvectors corresponding to distinct eigenvalues are linearly independent (see 5.10), this implies that each u_j equals 0. This implies (using 1.44) that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, as desired.

Now

$$\begin{aligned}\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) &= \dim\bigl(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)\bigr) \\ &\leq \dim V,\end{aligned}$$

where the equality above follows from Exercise 16 in Section 2.C. ■

5.39 Definition ***diagonalizable***

An operator $T \in \mathcal{L}(V)$ is called ***diagonalizable*** if the operator has a diagonal matrix with respect to some basis of V.

5.40 Example Define $T \in \mathcal{L}(\mathbf{R}^2)$ by

$$T(x, y) = (41x + 7y, -20x + 74y).$$

The matrix of T with respect to the standard basis of $\mathbf{R}^2$ is

$$\begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix},$$

which is not a diagonal matrix. However, T is diagonalizable, because the matrix of T with respect to the basis $(1, 4), (7, 5)$ is

$$\begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix},$$

as you should verify.

5.41 Conditions equivalent to diagonalizability

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

(a) T is diagonalizable;

(b) V has a basis consisting of eigenvectors of T;

(c) there exist 1-dimensional subspaces $U_1, \dots, U_n$ of V, each invariant under T, such that

$$V = U_1 \oplus \cdots \oplus U_n;$$

(d) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;

(e) $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Proof An operator $T \in \mathcal{L}(V)$ has a diagonal matrix

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

with respect to a basis $v_1, \dots, v_n$ of V if and only if $Tv_j = \lambda_j v_j$ for each j. Thus (a) and (b) are equivalent.

Suppose (b) holds; thus V has a basis $v_1, \dots, v_n$ consisting of eigenvectors of T. For each j, let $U_j = \operatorname{span}(v_j)$. Obviously each U_j is a 1-dimensional subspace of V that is invariant under T. Because $v_1, \dots, v_n$ is a basis of V, each vector in V can be written uniquely as a linear combination of $v_1, \dots, v_n$. In other words, each vector in V can be written uniquely as a sum $u_1 + \cdots + u_n$, where each u_j is in U_j. Thus $V = U_1 \oplus \cdots \oplus U_n$. Hence (b) implies (c).

Suppose now that (c) holds; thus there are 1-dimensional subspaces $U_1, \dots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus \cdots \oplus U_n$. For each j, let v_j be a nonzero vector in U_j. Then each v_j is an eigenvector of T. Because each vector in V can be written uniquely as a sum $u_1 + \cdots + u_n$, where each u_j is in U_j (so each u_j is a scalar multiple of v_j), we see that $v_1, \dots, v_n$ is a basis of V. Thus (c) implies (b).

At this stage of the proof we know that (a), (b), and (c) are all equivalent. We will finish the proof by showing that (b) implies (d), that (d) implies (e), and that (e) implies (b).

Suppose (b) holds; thus V has a basis consisting of eigenvectors of T. Hence every vector in V is a linear combination of eigenvectors of T, which implies that

$$V = E(\lambda_1, T) + \cdots + E(\lambda_m, T).$$

Now 5.38 shows that (d) holds.

That (d) implies (e) follows immediately from Exercise 16 in Section 2.C.

Finally, suppose (e) holds; thus

5.42 $$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T).$$

Choose a basis of each $E(\lambda_j, T)$; put all these bases together to form a list $v_1, \ldots, v_n$ of eigenvectors of T, where $n = \dim V$ (by 5.42). To show that this list is linearly independent, suppose

$$a_1 v_1 + \cdots + a_n v_n = 0,$$

where $a_1, \ldots, a_n \in \mathbf{F}$. For each $j = 1, \ldots, m$, let u_j denote the sum of all the terms $a_k v_k$ such that $v_k \in E(\lambda_j, T)$. Thus each u_j is in $E(\lambda_j, T)$, and

$$u_1 + \cdots + u_m = 0.$$

Because eigenvectors corresponding to distinct eigenvalues are linearly independent (see 5.10), this implies that each u_j equals 0. Because each u_j is a sum of terms $a_k v_k$, where the v_k's were chosen to be a basis of $E(\lambda_j, T)$, this implies that all the a_k's equal 0. Thus $v_1, \ldots, v_n$ is linearly independent and hence is a basis of V (by 2.39). Thus (e) implies (b), completing the proof. ■

Unfortunately not every operator is diagonalizable. This sad state of affairs can arise even on complex vector spaces, as shown by the next example.

5.43 **Example** Show that the operator $T \in \mathcal{L}(\mathbf{C}^2)$ defined by

$$T(w, z) = (z, 0)$$

is not diagonalizable.

Solution As you should verify, 0 is the only eigenvalue of T and furthermore $E(0, T) = \{(w, 0) \in \mathbf{C}^2 : w \in \mathbf{C}\}$.

Thus conditions (b), (c), (d), and (e) of 5.41 are easily seen to fail (of course, because these conditions are equivalent, it is only necessary to check that one of them fails). Thus condition (a) of 5.41 also fails, and hence T is not diagonalizable.

The next result shows that if an operator has as many distinct eigenvalues as the dimension of its domain, then the operator is diagonalizable.

5.44 Enough eigenvalues implies diagonalizability

If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.

Proof Suppose $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues $\lambda_1, \dots, \lambda_{\dim V}$. For each j, let $v_j \in V$ be an eigenvector corresponding to the eigenvalue λ_j. Because eigenvectors corresponding to distinct eigenvalues are linearly independent (see 5.10), $v_1, \dots, v_{\dim V}$ is linearly independent. A linearly independent list of $\dim V$ vectors in V is a basis of V (see 2.39); thus $v_1, \dots, v_{\dim V}$ is a basis of V. With respect to this basis consisting of eigenvectors, T has a diagonal matrix. ■

5.45 Example Define $T \in \mathcal{L}(\mathbf{F}^3)$ by $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. Find a basis of $\mathbf{F}^3$ with respect to which T has a diagonal matrix.

Solution With respect to the standard basis, the matrix of T is

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

The matrix above is upper triangular. Thus by 5.32, the eigenvalues of T are 2, 5, and 8. Because T is an operator on a vector space with dimension 3 and T has three distinct eigenvalues, 5.44 assures us that there exists a basis of $\mathbf{F}^3$ with respect to which T has a diagonal matrix.

To find this basis, we only have to find an eigenvector for each eigenvalue. In other words, we have to find a nonzero solution to the equation

$$T(x, y, z) = \lambda(x, y, z)$$

for $\lambda = 2$, then for $\lambda = 5$, and then for $\lambda = 8$. These simple equations are easy to solve: for $\lambda = 2$ we have the eigenvector $(1, 0, 0)$; for $\lambda = 5$ we have the eigenvector $(1, 3, 0)$; for $\lambda = 8$ we have the eigenvector $(1, 6, 6)$.

Thus $(1, 0, 0), (1, 3, 0), (1, 6, 6)$ is a basis of $\mathbf{F}^3$, and with respect to this basis the matrix of T is

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

The converse of 5.44 is not true. For example, the operator T defined on the three-dimensional space $\mathbf{F}^3$ by

$$T(z_1, z_2, z_3) = (4z_1, 4z_2, 5z_3)$$

has only two eigenvalues (4 and 5), but this operator has a diagonal matrix with respect to the standard basis.

In later chapters we will find additional conditions that imply that certain operators are diagonalizable.

EXERCISES 5.C

1 Suppose $T \in \mathcal{L}(V)$ is diagonalizable. Prove that $V = \operatorname{null} T \oplus \operatorname{range} T$.

dim V < +inf

2 Prove the converse of the statement in the exercise above or give a counterexample to the converse. (x, y) -> (x, -y)

3 Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Prove that the following are equivalent:

(a) $V = \operatorname{null} T \oplus \operatorname{range} T$.

(b) $V = \operatorname{null} T + \operatorname{range} T$.

(c) $\operatorname{null} T \cap \operatorname{range} T = \{0\}$.

4 Give an example to show that the exercise above is false without the hypothesis that V is finite-dimensional.

5 Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Prove that T is diagonalizable if and only if

$$V = \operatorname{null}(T - \lambda I) \oplus \operatorname{range}(T - \lambda I)$$

for every $\lambda \in \mathbf{C}$.

6 Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, and $S \in \mathcal{L}(V)$ has the same eigenvectors as T (not necessarily with the same eigenvalues). Prove that $ST = TS$.

7 Suppose $T \in \mathcal{L}(V)$ has a diagonal matrix A with respect to some basis of V and that $\lambda \in \mathbf{F}$. Prove that λ appears on the diagonal of A precisely $\dim E(\lambda, T)$ times.

8 Suppose $T \in \mathcal{L}(\mathbf{F}^5)$ and $\dim E(8, T) = 4$. Prove that $T - 2I$ or $T - 6I$ is invertible.

9 Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that $E(\lambda, T) = E(\frac{1}{\lambda}, T^{-1})$ for every $\lambda \in \mathbf{F}$ with $\lambda \neq 0$.

10 Suppose that V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct nonzero eigenvalues of T. Prove that

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim \operatorname{range} T.$$

11 Verify the assertion in Example 5.40.

12 Suppose $R, T \in \mathcal{L}(\mathbf{F}^3)$ each have 2, 6, 7 as eigenvalues. Prove that there exists an invertible operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $R = S^{-1}TS$.

13 Find $R, T \in \mathcal{L}(\mathbf{F}^4)$ such that R and T each have 2, 6, 7 as eigenvalues, R and T have no other eigenvalues, and there does not exist an invertible operator $S \in \mathcal{L}(\mathbf{F}^4)$ such that $R = S^{-1}TS$.

14 Find $T \in \mathcal{L}(\mathbf{C}^3)$ such that 6 and 7 are eigenvalues of T and such that T does not have a diagonal matrix with respect to any basis of $\mathbf{C}^3$.

15 Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is such that 6 and 7 are eigenvalues of T. Furthermore, suppose T does not have a diagonal matrix with respect to any basis of $\mathbf{C}^3$. Prove that there exists $(x, y, z) \in \mathbf{F}^3$ such that $T(x, y, z) = (17 + 8x, \sqrt{5} + 8y, 2\pi + 8z)$.

16 The ***Fibonacci sequence*** $F_1, F_2, \dots$ is defined by

$$F_1 = 1,\ F_2 = 1, \quad \text{and} \quad F_n = F_{n-2} + F_{n-1} \text{ for } n \geq 3.$$

Define $T \in \mathcal{L}(\mathbf{R}^2)$ by $T(x, y) = (y, x + y)$.

(a) Show that $T^n(0, 1) = (F_n, F_{n+1})$ for each positive integer n.

(b) Find the eigenvalues of T.

(c) Find a basis of $\mathbf{R}^2$ consisting of eigenvectors of T.

(d) Use the solution to part (c) to compute $T^n(0, 1)$. Conclude that

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

for each positive integer n.

(e) Use part (d) to conclude that for each positive integer n, the Fibonacci number F_n is the integer that is closest to

$$\frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n.$$

CHAPTER 6

Woman teaching geometry, from a fourteenth-century edition of Euclid's geometry book.

Inner Product Spaces

In making the definition of a vector space, we generalized the linear structure (addition and scalar multiplication) of $\mathbf{R}^2$ and $\mathbf{R}^3$. We ignored other important features, such as the notions of length and angle. These ideas are embedded in the concept we now investigate, inner products.

Our standing assumptions are as follows:

6.1 Notation $\mathbf{F}, V$

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a vector space over $\mathbf{F}$.

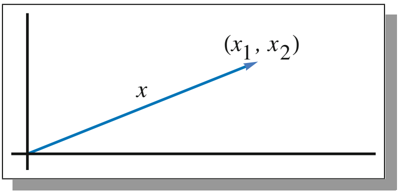
LEARNING OBJECTIVES FOR THIS CHAPTER

- Cauchy–Schwarz Inequality
- Gram–Schmidt Procedure
- linear functionals on inner product spaces
- calculating minimum distance to a subspace

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_6

6.A *Inner Products and Norms*

Inner Products



The length of this vector x is $\sqrt{x_1{}^2 + x_2{}^2}$.

To motivate the concept of inner product, think of vectors in $\mathbf{R}^2$ and $\mathbf{R}^3$ as arrows with initial point at the origin. The length of a vector x in $\mathbf{R}^2$ or $\mathbf{R}^3$ is called the ***norm*** of x, denoted $\|x\|$. Thus for $x = (x_1, x_2) \in \mathbf{R}^2$, we have $\|x\| = \sqrt{x_1{}^2 + x_2{}^2}$.

Similarly, if $x = (x_1, x_2, x_3) \in \mathbf{R}^3$, then $\|x\| = \sqrt{x_1{}^2 + x_2{}^2 + x_3{}^2}$.

Even though we cannot draw pictures in higher dimensions, the generalization to $\mathbf{R}^n$ is obvious: we define the norm of $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ by

$$\|x\| = \sqrt{x_1{}^2 + \cdots + x_n{}^2}.$$

The norm is not linear on $\mathbf{R}^n$. To inject linearity into the discussion, we introduce the dot product.

6.2 Definition ***dot product***

For $x, y \in \mathbf{R}^n$, the ***dot product*** of x and y, denoted $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$.

If we think of vectors as points instead of arrows, then $\|x\|$ should be interpreted as the distance from the origin to the point x.

Note that the dot product of two vectors in $\mathbf{R}^n$ is a number, not a vector. Obviously $x \cdot x = \|x\|^2$ for all $x \in \mathbf{R}^n$. The dot product on $\mathbf{R}^n$ has the following properties:

- $x \cdot x \geq 0$ for all $x \in \mathbf{R}^n$;
- $x \cdot x = 0$ if and only if $x = 0$;
- for $y \in \mathbf{R}^n$ fixed, the map from $\mathbf{R}^n$ to $\mathbf{R}$ that sends $x \in \mathbf{R}^n$ to $x \cdot y$ is linear;
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbf{R}^n$.

An inner product is a generalization of the dot product. At this point you may be tempted to guess that an inner product is defined by abstracting the properties of the dot product discussed in the last paragraph. For real vector spaces, that guess is correct. However, so that we can make a definition that will be useful for both real and complex vector spaces, we need to examine the complex case before making the definition.

Recall that if $\lambda = a + bi$, where $a, b \in \mathbf{R}$, then

- the absolute value of λ, denoted $|\lambda|$, is defined by $|\lambda| = \sqrt{a^2 + b^2}$;
- the complex conjugate of λ, denoted $\bar{\lambda}$, is defined by $\bar{\lambda} = a - bi$;
- $|\lambda|^2 = \lambda\bar{\lambda}$.

See Chapter 4 for the definitions and the basic properties of the absolute value and complex conjugate.

For $z = (z_1, \dots, z_n) \in \mathbf{C}^n$, we define the norm of z by

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

The absolute values are needed because we want $\|z\|$ to be a nonnegative number. Note that

$$\|z\|^2 = z_1\overline{z_1} + \cdots + z_n\overline{z_n}.$$

We want to think of $\|z\|^2$ as the inner product of z with itself, as we did in $\mathbf{R}^n$. The equation above thus suggests that the inner product of $w = (w_1, \dots, w_n) \in \mathbf{C}^n$ with z should equal

$$w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

If the roles of the w and z were interchanged, the expression above would be replaced with its complex conjugate. In other words, we should expect that the inner product of w with z equals the complex conjugate of the inner product of z with w. With that motivation, we are now ready to define an inner product on V, which may be a real or a complex vector space.

Two comments about the notation used in the next definition:

- If λ is a complex number, then the notation $\lambda \geq 0$ means that λ is real and nonnegative.
- We use the common notation $\langle u, v\rangle$, with angle brackets denoting an inner product. Some people use parentheses instead, but then (u, v) becomes ambiguous because it could denote either an ordered pair or an inner product.

6.3 Definition ***inner product***

An ***inner product*** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbf{F}$ and has the following properties:

positivity
$\langle v, v\rangle \geq 0$ for all $v \in V$;

definiteness
$\langle v, v\rangle = 0$ if and only if $v = 0$;

additivity in first slot
$\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$;

homogeneity in first slot
$\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ for all $\lambda \in \mathbf{F}$ and all $u, v \in V$;

conjugate symmetry
$\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$.

Although most mathematicians define an inner product as above, many physicists use a definition that requires homogeneity in the second slot instead of the first slot.

Every real number equals its complex conjugate. Thus if we are dealing with a real vector space, then in the last condition above we can dispense with the complex conjugate and simply state that $\langle u, v\rangle = \langle v, u\rangle$ for all $v, w \in V$.

6.4 Example ***inner products***

(a) The ***Euclidean inner product*** on $\mathbf{F}^n$ is defined by

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n)\rangle = w_1\overline{z_1} + \dots + w_n\overline{z_n}.$$

(b) If $c_1, \dots, c_n$ are positive numbers, then an inner product can be defined on $\mathbf{F}^n$ by

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n)\rangle = c_1 w_1\overline{z_1} + \dots + c_n w_n\overline{z_n}.$$

(c) An inner product can be defined on the vector space of continuous real-valued functions on the interval $[-1, 1]$ by

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx.$$

(d) An inner product can be defined on $\mathcal{P}(\mathbf{R})$ by

$$\langle p, q\rangle = \int_{0}^{\infty} p(x)q(x)e^{-x}\,dx.$$

6.5 Definition ***inner product space***

An ***inner product space*** is a vector space V along with an inner product on V.

The most important example of an inner product space is $\mathbf{F}^n$ with the Euclidean inner product given by part (a) of the last example. When $\mathbf{F}^n$ is referred to as an inner product space, you should assume that the inner product is the Euclidean inner product unless explicitly told otherwise.

So that we do not have to keep repeating the hypothesis that V is an inner product space, for the rest of this chapter we make the following assumption:

6.6 Notation V

For the rest of this chapter, V denotes an inner product space over $\mathbf{F}$.

Note the slight abuse of language here. An inner product space is a vector space along with an inner product on that vector space. When we say that a vector space V is an inner product space, we are also thinking that an inner product on V is lurking nearby or is obvious from the context (or is the Euclidean inner product if the vector space is $\mathbf{F}^n$).

6.7 Basic properties of an inner product

(a) For each fixed $u \in V$, the function that takes v to $\langle v, u\rangle$ is a linear map from V to $\mathbf{F}$.

(b) $\langle 0, u\rangle = 0$ for every $u \in V$.

(c) $\langle u, 0\rangle = 0$ for every $u \in V$.

(d) $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ for all $u, v, w \in V$.

(e) $\langle u, \lambda v\rangle = \bar{\lambda}\langle u, v\rangle$ for all $\lambda \in \mathbf{F}$ and $u, v \in V$.

Proof

(a) Part (a) follows from the conditions of additivity in the first slot and homogeneity in the first slot in the definition of an inner product.

(b) Part (b) follows from part (a) and the result that every linear map takes 0 to 0.

(c) Part (c) follows from part (a) and the conjugate symmetry property in the definition of an inner product.

(d) Suppose $u, v, w \in V$. Then

$$\begin{aligned}\langle u, v + w\rangle &= \overline{\langle v + w, u\rangle}\\ &= \overline{\langle v, u\rangle + \langle w, u\rangle}\\ &= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle}\\ &= \langle u, v\rangle + \langle u, w\rangle.\end{aligned}$$

(e) Suppose $\lambda \in \mathbf{F}$ and $u, v \in V$. Then

$$\begin{aligned}\langle u, \lambda v\rangle &= \overline{\langle \lambda v, u\rangle}\\ &= \overline{\lambda\langle v, u\rangle}\\ &= \bar{\lambda}\overline{\langle v, u\rangle}\\ &= \bar{\lambda}\langle u, v\rangle,\end{aligned}$$

as desired. ■

Norms

Our motivation for defining inner products came initially from the norms of vectors on $\mathbf{R}^2$ and $\mathbf{R}^3$. Now we see that each inner product determines a norm.

6.8 Definition ***norm,*** $\|v\|$

For $v \in V$, the ***norm*** of v, denoted $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

6.9 Example ***norms***

(a) If $(z_1, \ldots, z_n) \in \mathbf{F}^n$ (with the Euclidean inner product), then

$$\|(z_1, \ldots, z_n)\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

(b) In the vector space of continuous real-valued functions on $[-1, 1]$ [with inner product given as in part (c) of 6.4], we have

$$\|f\| = \sqrt{\int_{-1}^{1} \bigl(f(x)\bigr)^2\,dx}.$$

6.10 Basic properties of the norm

Suppose $v \in V$.

(a) $\|v\| = 0$ if and only if $v = 0$.

(b) $\|\lambda v\| = |\lambda|\,\|v\|$ for all $\lambda \in \mathbf{F}$.

Proof

(a) The desired result holds because $\langle v, v\rangle = 0$ if and only if $v = 0$.

(b) Suppose $\lambda \in \mathbf{F}$. Then

$$\begin{aligned}\|\lambda v\|^2 &= \langle \lambda v, \lambda v\rangle \\ &= \lambda\langle v, \lambda v\rangle \\ &= \lambda\bar{\lambda}\langle v, v\rangle \\ &= |\lambda|^2\|v\|^2.\end{aligned}$$

Taking square roots now gives the desired equality. ■

The proof above of part (b) illustrates a general principle: working with norms squared is usually easier than working directly with norms.

Now we come to a crucial definition.

6.11 Definition *orthogonal*

Two vectors $u, v \in V$ are called ***orthogonal*** if $\langle u, v\rangle = 0$.

In the definition above, the order of the vectors does not matter, because $\langle u, v\rangle = 0$ if and only if $\langle v, u\rangle = 0$. Instead of saying that u and v are orthogonal, sometimes we say that u is orthogonal to v.

Exercise 13 asks you to prove that if u, v are nonzero vectors in $\mathbf{R}^2$, then

$$\langle u, v\rangle = \|u\|\,\|v\|\cos\theta,$$

where θ is the angle between u and v (thinking of u and v as arrows with initial point at the origin). Thus two vectors in $\mathbf{R}^2$ are orthogonal (with respect to the usual Euclidean inner product) if and only if the cosine of the angle between them is 0, which happens if and only if the vectors are perpendicular in the usual sense of plane geometry. Thus you can think of the word *orthogonal* as a fancy word meaning *perpendicular*.

We begin our study of orthogonality with an easy result.

6.12 Orthogonality and 0

(a) 0 is orthogonal to every vector in V.

(b) 0 is the only vector in V that is orthogonal to itself.

Proof

(a) Part (b) of 6.7 states that $\langle 0, u \rangle = 0$ for every $u \in V$.

(b) If $v \in V$ and $\langle v, v \rangle = 0$, then $v = 0$ (by definition of inner product). ■

*The word **orthogonal** comes from the Greek word **orthogonios**, which means right-angled.*

For the special case $V = \mathbf{R}^2$, the next theorem is over 2,500 years old. Of course, the proof below is not the original proof.

6.13 Pythagorean Theorem

Suppose u and v are orthogonal vectors in V. Then

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

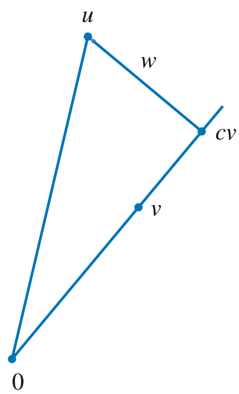
Proof We have

$$\begin{aligned} \|u + v\|^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= \|u\|^2 + \|v\|^2, \end{aligned}$$

as desired. ■

The proof given above of the Pythagorean Theorem shows that the conclusion holds if and only if $\langle u, v \rangle + \langle v, u \rangle$, which equals $2\,\mathrm{Re}\langle u, v \rangle$, is 0. Thus the converse of the Pythagorean Theorem holds in real inner product spaces.

Suppose $u, v \in V$, with $v \neq 0$. We would like to write u as a scalar multiple of v plus a vector w orthogonal to v, as suggested in the next picture.



An orthogonal decomposition.

To discover how to write u as a scalar multiple of v plus a vector orthogonal to v, let $c \in \mathbf{F}$ denote a scalar. Then

$$u = cv + (u - cv).$$

Thus we need to choose c so that v is orthogonal to $(u - cv)$. In other words, we want

$$0 = \langle u - cv, v \rangle = \langle u, v \rangle - c\|v\|^2.$$

The equation above shows that we should choose c to be $\langle u, v \rangle / \|v\|^2$. Making this choice of c, we can write

$$u = \frac{\langle u, v \rangle}{\|v\|^2} v + \left(u - \frac{\langle u, v \rangle}{\|v\|^2} v \right).$$

As you should verify, the equation above writes u as a scalar multiple of v plus a vector orthogonal to v. In other words, we have proved the following result.

6.14 An orthogonal decomposition

Suppose $u, v \in V$, with $v \neq 0$. Set $c = \dfrac{\langle u, v \rangle}{\|v\|^2}$ and $w = u - \dfrac{\langle u, v \rangle}{\|v\|^2} v$. Then

$$\langle w, v \rangle = 0 \quad \text{and} \quad u = cv + w.$$

The orthogonal decomposition 6.14 will be used in the proof of the Cauchy–Schwarz Inequality, which is our next result and is one of the most important inequalities in mathematics.

French mathematician Augustin-Louis Cauchy (1789–1857) proved 6.17(a) in 1821. German mathematician Hermann Schwarz (1843–1921) proved 6.17(b) in 1886.

6.15 **Cauchy–Schwarz Inequality**

Suppose $u, v \in V$. Then

$$|\langle u, v\rangle| \leq \|u\| \, \|v\|.$$

This inequality is an equality if and only if one of u, v is a scalar multiple of the other.

Proof If $v = 0$, then both sides of the desired inequality equal 0. Thus we can assume that $v \neq 0$. Consider the orthogonal decomposition

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + w$$

given by 6.14, where w is orthogonal to v. By the Pythagorean Theorem,

$$\begin{aligned} \|u\|^2 &= \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\|^2 + \|w\|^2 \\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\ &\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}. \qquad \textbf{6.16} \end{aligned}$$

Multiplying both sides of this inequality by $\|v\|^2$ and then taking square roots gives the desired inequality.

Looking at the proof in the paragraph above, note that the Cauchy–Schwarz Inequality is an equality if and only if 6.16 is an equality. Obviously this happens if and only if $w = 0$. But $w = 0$ if and only if u is a multiple of v (see 6.14). Thus the Cauchy–Schwarz Inequality is an equality if and only if u is a scalar multiple of v or v is a scalar multiple of u (or both; the phrasing has been chosen to cover cases in which either u or v equals 0). ■

6.17 **Example** ***examples of the Cauchy–Schwarz Inequality***

(a) If $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbf{R}$, then

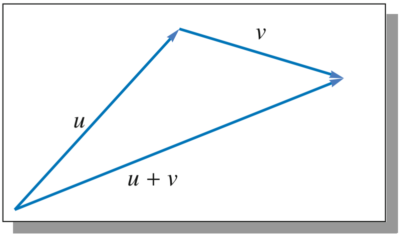
$$|x_1 y_1 + \dots + x_n y_n|^2 \leq ({x_1}^2 + \dots + {x_n}^2)({y_1}^2 + \dots + {y_n}^2).$$

(b) If f, g are continuous real-valued functions on $[-1, 1]$, then

$$\left| \int_{-1}^{1} f(x) g(x) \, dx \right|^2 \leq \left(\int_{-1}^{1} \bigl(f(x)\bigr)^2 \, dx \right) \left(\int_{-1}^{1} \bigl(g(x)\bigr)^2 \, dx \right).$$

The next result, called the Triangle Inequality, has the geometric interpretation that the length of each side of a triangle is less than the sum of the lengths of the other two sides.

Note that the Triangle Inequality implies that the shortest path between two points is a line segment.



6.18 Triangle Inequality

Suppose $u, v \in V$. Then

$$\|u + v\| \le \|u\| + \|v\|.$$

This inequality is an equality if and only if one of u, v is a nonnegative multiple of the other.

Proof We have

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\operatorname{Re}\langle u, v\rangle \\
&\le \|u\|^2 + \|v\|^2 + 2|\langle u, v\rangle| \qquad (6.19) \\
&\le \|u\|^2 + \|v\|^2 + 2\|u\|\,\|v\| \qquad (6.20) \\
&= (\|u\| + \|v\|)^2,
\end{aligned}$$

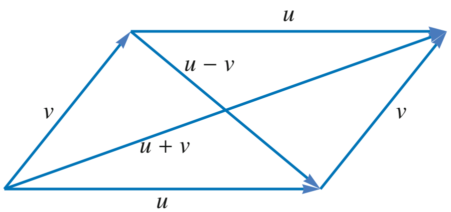
where 6.20 follows from the Cauchy–Schwarz Inequality (6.15). Taking square roots of both sides of the inequality above gives the desired inequality.

The proof above shows that the Triangle Inequality is an equality if and only if we have equality in 6.19 and 6.20. Thus we have equality in the Triangle Inequality if and only if

$$\langle u, v\rangle = \|u\|\|v\|. \qquad (6.21)$$

If one of u, v is a nonnegative multiple of the other, then 6.21 holds, as you should verify. Conversely, suppose 6.21 holds. Then the condition for equality in the Cauchy–Schwarz Inequality (6.15) implies that one of u, v is a scalar multiple of the other. Clearly 6.21 forces the scalar in question to be nonnegative, as desired. ■

The next result is called the parallelogram equality because of its geometric interpretation: in every parallelogram, the sum of the squares of the lengths of the diagonals equals the sum of the squares of the lengths of the four sides.



The parallelogram equality.

6.22 Parallelogram Equality

Suppose $u, v \in V$. Then

$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Proof We have

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\ &\qquad + \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\ &= 2(\|u\|^2 + \|v\|^2),\end{aligned}$$

as desired. ■

Law professor Richard Friedman presenting a case before the U.S. Supreme Court in 2010:

Mr. Friedman: I think that issue is entirely orthogonal to the issue here because the Commonwealth is acknowledging—
Chief Justice Roberts: I'm sorry. Entirely what?
Mr. Friedman: Orthogonal. Right angle. Unrelated. Irrelevant.
Chief Justice Roberts: Oh.
Justice Scalia: What was that adjective? I liked that.
Mr. Friedman: Orthogonal.
Chief Justice Roberts: Orthogonal.
Mr. Friedman: Right, right.
Justice Scalia: Orthogonal, ooh. (Laughter.)
Justice Kennedy: I knew this case presented us a problem. (Laughter.)

EXERCISES 6.A

1 Show that the function that takes $\big((x_1, x_2), (y_1, y_2)\big) \in \mathbf{R}^2 \times \mathbf{R}^2$ to $|x_1 y_1| + |x_2 y_2|$ is not an inner product on $\mathbf{R}^2$.

2 Show that the function that takes $\big((x_1, x_2, x_3), (y_1, y_2, y_3)\big) \in \mathbf{R}^3 \times \mathbf{R}^3$ to $x_1 y_1 + x_3 y_3$ is not an inner product on $\mathbf{R}^3$.

3 Suppose $\mathbf{F} = \mathbf{R}$ and $V \neq \{0\}$. Replace the positivity condition (which states that $\langle v, v\rangle \geq 0$ for all $v \in V$) in the definition of an inner product (6.3) with the condition that $\langle v, v\rangle > 0$ for some $v \in V$. Show that this change in the definition does not change the set of functions from $V \times V$ to $\mathbf{R}$ that are inner products on V.

4 Suppose V is a real inner product space.

(a) Show that $\langle u + v, u - v\rangle = \|u\|^2 - \|v\|^2$ for every $u, v \in V$.

(b) Show that if $u, v \in V$ have the same norm, then $u + v$ is orthogonal to $u - v$.

(c) Use part (b) to show that the diagonals of a rhombus are perpendicular to each other.

5 Suppose $T \in \mathcal{L}(V)$ is such that $\|Tv\| \leq \|v\|$ for every $v \in V$. Prove that $T - \sqrt{2}I$ is invertible.

6 Suppose $u, v \in V$. Prove that $\langle u, v\rangle = 0$ if and only if

$$\|u\| \leq \|u + av\|$$

for all $a \in \mathbf{F}$.

7 Suppose $u, v \in V$. Prove that $\|au + bv\| = \|bu + av\|$ for all $a, b \in \mathbf{R}$ if and only if $\|u\| = \|v\|$.

8 Suppose $u, v \in V$ and $\|u\| = \|v\| = 1$ and $\langle u, v\rangle = 1$. Prove that $u = v$.

9 Suppose $u, v \in V$ and $\|u\| \leq 1$ and $\|v\| \leq 1$. Prove that

$$\sqrt{1 - \|u\|^2}\sqrt{1 - \|v\|^2} \leq 1 - |\langle u, v\rangle|.$$

10 Find vectors $u, v \in \mathbf{R}^2$ such that u is a scalar multiple of $(1, 3)$, v is orthogonal to $(1, 3)$, and $(1, 2) = u + v$.

11 Prove that

$$16 \leq (a+b+c+d)\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right)$$

for all positive numbers a, b, c, d.

12 Prove that

$$(x_1 + \cdots + x_n)^2 \leq n({x_1}^2 + \cdots + {x_n}^2)$$

for all positive integers n and all real numbers $x_1, \ldots, x_n$.

13 Suppose u, v are nonzero vectors in $\mathbf{R}^2$. Prove that

$$\langle u, v \rangle = \|u\| \|v\| \cos\theta,$$

where θ is the angle between u and v (thinking of u and v as arrows with initial point at the origin).

Hint: Draw the triangle formed by u, v, and $u - v$; then use the law of cosines.

14 The angle between two vectors (thought of as arrows with initial point at the origin) in $\mathbf{R}^2$ or $\mathbf{R}^3$ can be defined geometrically. However, geometry is not as clear in $\mathbf{R}^n$ for $n > 3$. Thus the angle between two nonzero vectors $x, y \in \mathbf{R}^n$ is defined to be

$$\arccos \frac{\langle x, y \rangle}{\|x\| \|y\|},$$

where the motivation for this definition comes from the previous exercise. Explain why the Cauchy–Schwarz Inequality is needed to show that this definition makes sense.

15 Prove that

$$\Big(\sum_{j=1}^{n} a_j b_j\Big)^2 \leq \Big(\sum_{j=1}^{n} j {a_j}^2\Big)\Big(\sum_{j=1}^{n} \frac{{b_j}^2}{j}\Big)$$

for all real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$.

16 Suppose $u, v \in V$ are such that

$$\|u\| = 3, \quad \|u + v\| = 4, \quad \|u - v\| = 6.$$

What number does $\|v\|$ equal?

17 Prove or disprove: there is an inner product on $\mathbf{R}^2$ such that the associated norm is given by

$$\|(x, y)\| = \max\{x, y\}$$

for all $(x, y) \in \mathbf{R}^2$.

18 Suppose $p > 0$. Prove that there is an inner product on $\mathbf{R}^2$ such that the associated norm is given by

$$\|(x, y)\| = (x^p + y^p)^{1/p}$$

for all $(x, y) \in \mathbf{R}^2$ if and only if $p = 2$.

19 Suppose V is a real inner product space. Prove that

$$\langle u, v \rangle = \frac{\|u + v\|^2 - \|u - v\|^2}{4}$$

for all $u, v \in V$.

20 Suppose V is a complex inner product space. Prove that

$$\langle u, v \rangle = \frac{\|u + v\|^2 - \|u - v\|^2 + \|u + iv\|^2 i - \|u - iv\|^2 i}{4}$$

for all $u, v \in V$.

21 A norm on a vector space U is a function $\| \ \| \colon U \to [0, \infty)$ such that $\|u\| = 0$ if and only if $u = 0$, $\|\alpha u\| = |\alpha| \|u\|$ for all $\alpha \in \mathbf{F}$ and all $u \in U$, and $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in U$. Prove that a norm satisfying the parallelogram equality comes from an inner product (in other words, show that if $\| \ \|$ is a norm on U satisfying the parallelogram equality, then there is an inner product $\langle \ , \ \rangle$ on U such that $\|u\| = \langle u, u \rangle^{1/2}$ for all $u \in U$).

22 Show that the square of an average is less than or equal to the average of the squares. More precisely, show that if $a_1, \dots, a_n \in \mathbf{R}$, then the square of the average of $a_1, \dots, a_n$ is less than or equal to the average of ${a_1}^2, \dots, {a_n}^2$.

23 Suppose $V_1, \dots, V_m$ are inner product spaces. Show that the equation

$$\langle (u_1, \dots, u_m), (v_1, \dots, v_m) \rangle = \langle u_1, v_1 \rangle + \cdots + \langle u_m, v_m \rangle$$

defines an inner product on $V_1 \times \cdots \times V_m$.
[*In the expression above on the right, $\langle u_1, v_1 \rangle$ denotes the inner product on $V_1, \dots, \langle u_m, v_m \rangle$ denotes the inner product on V_m. Each of the spaces $V_1, \dots, V_m$ may have a different inner product, even though the same notation is used here.*]

24 Suppose $S \in \mathcal{L}(V)$ is an injective operator on V. Define $\langle \cdot, \cdot \rangle_1$ by

$$\langle u, v \rangle_1 = \langle Su, Sv \rangle$$

for $u, v \in V$. Show that $\langle \cdot, \cdot \rangle_1$ is an inner product on V.

25 Suppose $S \in \mathcal{L}(V)$ is not injective. Define $\langle \cdot, \cdot \rangle_1$ as in the exercise above. Explain why $\langle \cdot, \cdot \rangle_1$ is not an inner product on V.

26 Suppose f, g are differentiable functions from $\mathbf{R}$ to $\mathbf{R}^n$.

(a) Show that

$$\langle f(t), g(t) \rangle' = \langle f'(t), g(t) \rangle + \langle f(t), g'(t) \rangle.$$

(b) Suppose $c > 0$ and $\|f(t)\| = c$ for every $t \in \mathbf{R}$. Show that $\langle f'(t), f(t) \rangle = 0$ for every $t \in \mathbf{R}$.

(c) Interpret the result in part (b) geometrically in terms of the tangent vector to a curve lying on a sphere in $\mathbf{R}^n$ centered at the origin.

[*For the exercise above, a function $f : \mathbf{R} \to \mathbf{R}^n$ is called differentiable if there exist differentiable functions $f_1, \ldots, f_n$ from $\mathbf{R}$ to $\mathbf{R}$ such that $f(t) = \big(f_1(t), \ldots, f_n(t)\big)$ for each $t \in \mathbf{R}$. Furthermore, for each $t \in \mathbf{R}$, the derivative $f'(t) \in \mathbf{R}^n$ is defined by $f'(t) = \big({f_1}'(t), \ldots, {f_n}'(t)\big)$.*]

27 Suppose $u, v, w \in V$. Prove that

$$\|w - \tfrac{1}{2}(u + v)\|^2 = \frac{\|w - u\|^2 + \|w - v\|^2}{2} - \frac{\|u - v\|^2}{4}.$$

28 Suppose C is a subset of V with the property that $u, v \in C$ implies $\frac{1}{2}(u + v) \in C$. Let $w \in V$. Show that there is at most one point in C that is closest to w. In other words, show that there is at most one $u \in C$ such that

$$\|w - u\| \le \|w - v\| \quad \text{for all } v \in C.$$

Hint: Use the previous exercise.

29 For $u, v \in V$, define $d(u, v) = \|u - v\|$.

(a) Show that d is a metric on V.

(b) Show that if V is finite-dimensional, then d is a complete metric on V (meaning that every Cauchy sequence converges).

(c) Show that every finite-dimensional subspace of V is a closed subset of V (with respect to the metric d).

30 Fix a positive integer n. The ***Laplacian*** Δp of a twice differentiable function p on $\mathbf{R}^n$ is the function on $\mathbf{R}^n$ defined by

$$\Delta p = \frac{\partial^2 p}{\partial x_1^2} + \cdots + \frac{\partial^2 p}{\partial x_n^2}.$$

The function p is called ***harmonic*** if $\Delta p = 0$.

A ***polynomial*** on $\mathbf{R}^n$ is a linear combination of functions of the form ${x_1}^{m_1} \cdots {x_n}^{m_n}$, where $m_1, \dots, m_n$ are nonnegative integers.

Suppose q is a polynomial on $\mathbf{R}^n$. Prove that there exists a harmonic polynomial p on $\mathbf{R}^n$ such that $p(x) = q(x)$ for every $x \in \mathbf{R}^n$ with $\|x\| = 1$.

[*The only fact about harmonic functions that you need for this exercise is that if p is a harmonic function on $\mathbf{R}^n$ and $p(x) = 0$ for all $x \in \mathbf{R}^n$ with $\|x\| = 1$, then $p = 0$.*]

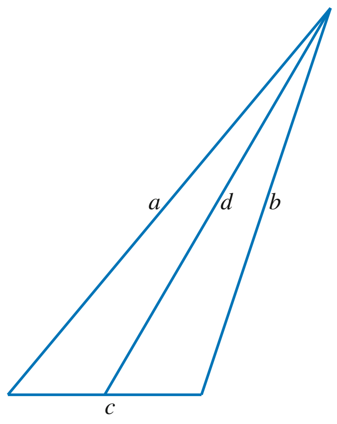
Hint: A reasonable guess is that the desired harmonic polynomial p is of the form $q + (1 - \|x\|^2)r$ for some polynomial r. Prove that there is a polynomial r on $\mathbf{R}^n$ such that $q + (1 - \|x\|^2)r$ is harmonic by defining an operator T on a suitable vector space by

$$Tr = \Delta\big((1 - \|x\|^2)r\big)$$

and then showing that T is injective and hence surjective.

31 Use inner products to prove Apollonius's Identity: In a triangle with sides of length a, b, and c, let d be the length of the line segment from the midpoint of the side of length c to the opposite vertex. Then

$$a^2 + b^2 = \tfrac{1}{2}c^2 + 2d^2.$$



6.B Orthonormal Bases

6.23 Definition ***orthonormal***

- A list of vectors is called ***orthonormal*** if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.
- In other words, a list $e_1, \ldots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

6.24 Example ***orthonormal lists***

(a) The standard basis in $\mathbf{F}^n$ is an orthonormal list.

(b) $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ is an orthonormal list in $\mathbf{F}^3$.

(c) $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)$ is an orthonormal list in $\mathbf{F}^3$.

Orthonormal lists are particularly easy to work with, as illustrated by the next result.

6.25 The norm of an orthonormal linear combination

If $e_1, \ldots, e_m$ is an orthonormal list of vectors in V, then

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \ldots, a_m \in \mathbf{F}$.

Proof Because each e_j has norm 1, this follows easily from repeated applications of the Pythagorean Theorem (6.13). ■

The result above has the following important corollary.

6.26 An orthonormal list is linearly independent

Every orthonormal list of vectors is linearly independent.

Proof Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in V and $a_1, \dots, a_m \in \mathbf{F}$ are such that

$$a_1 e_1 + \cdots + a_m e_m = 0.$$

Then $|a_1|^2 + \cdots + |a_m|^2 = 0$ (by 6.25), which means that all the a_j's are 0. Thus $e_1, \dots, e_m$ is linearly independent. ■

6.27 Definition ***orthonormal basis***

An ***orthonormal basis*** of V is an orthonormal list of vectors in V that is also a basis of V.

For example, the standard basis is an orthonormal basis of $\mathbf{F}^n$.

6.28 An orthonormal list of the right length is an orthonormal basis

Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.

Proof By 6.26, any such list must be linearly independent; because it has the right length, it is a basis—see 2.39. ■

6.29 Example Show that

$$\left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right), \left(\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}\right), \left(\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}\right), \left(-\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}\right)$$

is an orthonormal basis of $\mathbf{F}^4$.

Solution We have

$$\left\| \left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right) \right\| = \sqrt{\left(\tfrac{1}{2}\right)^2 + \left(\tfrac{1}{2}\right)^2 + \left(\tfrac{1}{2}\right)^2 + \left(\tfrac{1}{2}\right)^2} = 1.$$

Similarly, the other three vectors in the list above also have norm 1.

We have

$$\left\langle \left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right), \left(\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}\right) \right\rangle = \tfrac{1}{2} \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot \tfrac{1}{2} + \tfrac{1}{2} \cdot \left(-\tfrac{1}{2}\right) + \tfrac{1}{2} \cdot \left(-\tfrac{1}{2}\right) = 0.$$

Similarly, the inner product of any two distinct vectors in the list above also equals 0.

Thus the list above is orthonormal. Because we have an orthonormal list of length four in the four-dimensional vector space $\mathbf{F}^4$, this list is an orthonormal basis of $\mathbf{F}^4$ (by 6.28).

In general, given a basis $e_1, \dots, e_n$ of V and a vector $v \in V$, we know that there is some choice of scalars $a_1, \dots, a_n \in \mathbf{F}$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

The importance of orthonormal bases stems mainly from the next result.

Computing the numbers $a_1, \dots, a_n$ that satisfy the equation above can be difficult for an arbitrary basis of V. The next result shows, however, that this is easy for an orthonormal basis—just take $a_j = \langle v, e_j \rangle$.

6.30 Writing a vector as linear combination of orthonormal basis

Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

Proof Because $e_1, \dots, e_n$ is a basis of V, there exist scalars $a_1, \dots, a_n$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Because $e_1, \dots, e_n$ is orthonormal, taking the inner product of both sides of this equation with e_j gives $\langle v, e_j \rangle = a_j$. Thus the first equation in 6.30 holds.

The second equation in 6.30 follows immediately from the first equation and 6.25. ■

Now that we understand the usefulness of orthonormal bases, how do we go about finding them? For example, does $\mathcal{P}_m(\mathbf{R})$, with inner product given by integration on $[-1, 1]$ [see 6.4(c)], have an orthonormal basis? The next result will lead to answers to these questions.

Danish mathematician Jørgen Gram (1850–1916) and German mathematician Erhard Schmidt (1876–1959) popularized this algorithm that constructs orthonormal lists.

The algorithm used in the next proof is called the ***Gram–Schmidt Procedure***. It gives a method for turning a linearly independent list into an orthonormal list with the same span as the original list.

6.31 Gram–Schmidt Procedure

Suppose $v_1, \dots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \dots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}.$$

Then $e_1, \dots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$$

for $j = 1, \dots, m$.

Proof We will show by induction on j that the desired conclusion holds. To get started with $j = 1$, note that $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ because v_1 is a positive multiple of e_1.

Suppose $1 < j < m$ and we have verified that

6.32 $$\operatorname{span}(v_1, \dots, v_{j-1}) = \operatorname{span}(e_1, \dots, e_{j-1}).$$

Note that $v_j \notin \operatorname{span}(v_1, \dots, v_{j-1})$ (because $v_1, \dots, v_m$ is linearly independent). Thus $v_j \notin \operatorname{span}(e_1, \dots, e_{j-1})$. Hence we are not dividing by 0 in the definition of e_j given in 6.31. Dividing a vector by its norm produces a new vector with norm 1; thus $\|e_j\| = 1$.

Let $1 \le k < j$. Then

$$\begin{aligned}
\langle e_j, e_k\rangle &= \left\langle \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}, e_k \right\rangle \\
&= \frac{\langle v_j, e_k\rangle - \langle v_j, e_k\rangle}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|} \\
&= 0.
\end{aligned}$$

Thus $e_1, \dots, e_j$ is an orthonormal list.

From the definition of e_j given in 6.31, we see that $v_j \in \operatorname{span}(e_1, \dots, e_j)$. Combining this information with 6.32 shows that

$$\operatorname{span}(v_1, \dots, v_j) \subset \operatorname{span}(e_1, \dots, e_j).$$

Both lists above are linearly independent (the v's by hypothesis, the e's by orthonormality and 6.26). Thus both subspaces above have dimension j, and hence they are equal, completing the proof. ■

6.33 Example Find an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$, where the inner product is given by $\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\,dx$.

Solution We will apply the Gram–Schmidt Procedure (6.31) to the basis $1, x, x^2$.

To get started, with this inner product we have

$$\|1\|^2 = \int_{-1}^{1} 1^2\,dx = 2.$$

Thus $\|1\| = \sqrt{2}$, and hence $e_1 = \sqrt{\frac{1}{2}}$.

Now the numerator in the expression for e_2 is

$$x - \langle x, e_1\rangle e_1 = x - \Big(\int_{-1}^{1} x\sqrt{\tfrac{1}{2}}\,dx\Big)\sqrt{\tfrac{1}{2}} = x.$$

We have

$$\|x\|^2 = \int_{-1}^{1} x^2\,dx = \tfrac{2}{3}.$$

Thus $\|x\| = \sqrt{\frac{2}{3}}$, and hence $e_2 = \sqrt{\frac{3}{2}}x$.

Now the numerator in the expression for e_3 is

$$\begin{aligned} x^2 &- \langle x^2, e_1\rangle e_1 - \langle x^2, e_2\rangle e_2 \\ &= x^2 - \Big(\int_{-1}^{1} x^2\sqrt{\tfrac{1}{2}}\,dx\Big)\sqrt{\tfrac{1}{2}} - \Big(\int_{-1}^{1} x^2\sqrt{\tfrac{3}{2}}x\,dx\Big)\sqrt{\tfrac{3}{2}}x \\ &= x^2 - \tfrac{1}{3}. \end{aligned}$$

We have

$$\|x^2 - \tfrac{1}{3}\|^2 = \int_{-1}^{1} \big(x^4 - \tfrac{2}{3}x^2 + \tfrac{1}{9}\big)\,dx = \tfrac{8}{45}.$$

Thus $\|x^2 - \frac{1}{3}\| = \sqrt{\frac{8}{45}}$, and hence $e_3 = \sqrt{\frac{45}{8}}\big(x^2 - \frac{1}{3}\big)$.

Thus

$$\sqrt{\tfrac{1}{2}},\ \sqrt{\tfrac{3}{2}}x,\ \sqrt{\tfrac{45}{8}}\big(x^2 - \tfrac{1}{3}\big)$$

is an orthonormal list of length 3 in $\mathcal{P}_2(\mathbf{R})$. Hence this orthonormal list is an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$ by 6.28.

Now we can answer the question about the existence of orthonormal bases.

6.34 Existence of orthonormal basis

Every finite-dimensional inner product space has an orthonormal basis.

Proof Suppose V is finite-dimensional. Choose a basis of V. Apply the Gram–Schmidt Procedure (6.31) to it, producing an orthonormal list with length $\dim V$. By 6.28, this orthonormal list is an orthonormal basis of V. ■

Sometimes we need to know not only that an orthonormal basis exists, but also that every orthonormal list can be extended to an orthonormal basis. In the next corollary, the Gram–Schmidt Procedure shows that such an extension is always possible.

6.35 Orthonormal list extends to orthonormal basis

Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Proof Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in V. Then $e_1, \dots, e_m$ is linearly independent (by 6.26). Hence this list can be extended to a basis $e_1, \dots, e_m, v_1, \dots, v_n$ of V (see 2.33). Now apply the Gram–Schmidt Procedure (6.31) to $e_1, \dots, e_m, v_1, \dots, v_n$, producing an orthonormal list

6.36 $$e_1, \dots, e_m, f_1, \dots, f_n;$$

here the formula given by the Gram–Schmidt Procedure leaves the first m vectors unchanged because they are already orthonormal. The list above is an orthonormal basis of V by 6.28. ■

Recall that a matrix is called upper triangular if all entries below the diagonal equal 0. In other words, an upper-triangular matrix looks like this:

$$\begin{pmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{pmatrix},$$

where the 0 in the matrix above indicates that all entries below the diagonal equal 0, and asterisks are used to denote entries on and above the diagonal.

In the last chapter we showed that if V is a finite-dimensional complex vector space, then for each operator on V there is a basis with respect to which the matrix of the operator is upper triangular (see 5.27). Now that we are dealing with inner product spaces, we would like to know whether there exists an *orthonormal* basis with respect to which we have an upper-triangular matrix.

The next result shows that the existence of a basis with respect to which T has an upper-triangular matrix implies the existence of an orthonormal basis with this property. This result is true on both real and complex vector spaces (although on a real vector space, the hypothesis holds only for some operators).

6.37 Upper-triangular matrix with respect to orthonormal basis

Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof Suppose T has an upper-triangular matrix with respect to some basis $v_1, \dots, v_n$ of V. Thus $\operatorname{span}(v_1, \dots, v_j)$ is invariant under T for each $j = 1, \dots, n$ (see 5.26).

Apply the Gram–Schmidt Procedure to $v_1, \dots, v_n$, producing an orthonormal basis $e_1, \dots, e_n$ of V. Because

$$\operatorname{span}(e_1, \dots, e_j) = \operatorname{span}(v_1, \dots, v_j)$$

for each j (see 6.31), we conclude that $\operatorname{span}(e_1, \dots, e_j)$ is invariant under T for each $j = 1, \dots, n$. Thus, by 5.26, T has an upper-triangular matrix with respect to the orthonormal basis $e_1, \dots, e_n$. ■

German mathematician Issai Schur (1875–1941) published the first proof of the next result in 1909.

The next result is an important application of the result above.

6.38 Schur's Theorem

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Proof Recall that T has an upper-triangular matrix with respect to some basis of V (see 5.27). Now apply 6.37. ■

Linear Functionals on Inner Product Spaces

Because linear maps into the scalar field $\mathbf{F}$ play a special role, we defined a special name for them in Section 3.F. That definition is repeated below in case you skipped Section 3.F.

6.39 Definition ***linear functional***

A ***linear functional*** on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

6.40 Example The function $\varphi\colon \mathbf{F}^3 \to \mathbf{F}$ defined by

$$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

is a linear functional on $\mathbf{F}^3$. We could write this linear functional in the form

$$\varphi(z) = \langle z, u \rangle$$

for every $z \in \mathbf{F}^3$, where $u = (2, -5, 1)$.

6.41 Example The function $\varphi\colon \mathcal{P}_2(\mathbf{R}) \to \mathbf{R}$ defined by

$$\varphi(p) = \int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\, dt$$

is a linear functional on $\mathcal{P}_2(\mathbf{R})$ (here the inner product on $\mathcal{P}_2(\mathbf{R})$ is multiplication followed by integration on $[-1, 1]$; see 6.33). It is not obvious that there exists $u \in \mathcal{P}_2(\mathbf{R})$ such that

$$\varphi(p) = \langle p, u \rangle$$

for every $p \in \mathcal{P}_2(\mathbf{R})$ [we cannot take $u(t) = \cos(\pi t)$ because that function is not an element of $\mathcal{P}_2(\mathbf{R})$].

If $u \in V$, then the map that sends v to $\langle v, u \rangle$ is a linear functional on V. The next result shows that every linear functional on V is of this form. Example 6.41 above illustrates the power of the next result because for the linear functional in that example, there is no obvious candidate for u.

The next result is named in honor of Hungarian mathematician Frigyes Riesz (1880–1956), who proved several results early in the twentieth century that look very much like the result below.

6.42 Riesz Representation Theorem

Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

$$\varphi(v) = \langle v, u \rangle$$

for every $v \in V$.

Proof First we show there exists a vector $u \in V$ such that $\varphi(v) = \langle v, u \rangle$ for every $v \in V$. Let $e_1, \dots, e_n$ be an orthonormal basis of V. Then

$$\begin{aligned} \varphi(v) &= \varphi(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) \\ &= \langle v, e_1 \rangle \varphi(e_1) + \cdots + \langle v, e_n \rangle \varphi(e_n) \\ &= \langle v, \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n \rangle \end{aligned}$$

for every $v \in V$, where the first equality comes from 6.30. Thus setting

6.43 $$u = \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n,$$

we have $\varphi(v) = \langle v, u \rangle$ for every $v \in V$, as desired.

Now we prove that only one vector $u \in V$ has the desired behavior. Suppose $u_1, u_2 \in V$ are such that

$$\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle$$

for every $v \in V$. Then

$$0 = \langle v, u_1 \rangle - \langle v, u_2 \rangle = \langle v, u_1 - u_2 \rangle$$

for every $v \in V$. Taking $v = u_1 - u_2$ shows that $u_1 - u_2 = 0$. In other words, $u_1 = u_2$, completing the proof of the uniqueness part of the result. ■

6.44 **Example** Find $u \in \mathcal{P}_2(\mathbf{R})$ such that

$$\int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\,dt = \int_{-1}^{1} p(t)u(t)\,dt$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

Solution Let $\varphi(p) = \int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\,dt$. Applying formula 6.43 from the proof above, and using the orthonormal basis from Example 6.33, we have

$$u(x) = \Big(\int_{-1}^{1} \sqrt{\tfrac{1}{2}}\big(\cos(\pi t)\big)\,dt\Big)\sqrt{\tfrac{1}{2}} + \Big(\int_{-1}^{1} \sqrt{\tfrac{3}{2}}t\big(\cos(\pi t)\big)\,dt\Big)\sqrt{\tfrac{3}{2}}x$$
$$+ \Big(\int_{-1}^{1} \sqrt{\tfrac{45}{8}}\big(t^2 - \tfrac{1}{3}\big)\big(\cos(\pi t)\big)\,dt\Big)\sqrt{\tfrac{45}{8}}\big(x^2 - \tfrac{1}{3}\big).$$

A bit of calculus shows that

$$u(x) = -\tfrac{45}{2\pi^2}\big(x^2 - \tfrac{1}{3}\big).$$

Suppose V is finite-dimensional and φ a linear functional on V. Then 6.43 gives a formula for the vector u that satisfies $\varphi(v) = \langle v, u\rangle$ for all $v \in V$. Specifically, we have

$$u = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n.$$

The right side of the equation above seems to depend on the orthonormal basis $e_1, \dots, e_n$ as well as on φ. However, 6.42 tells us that u is uniquely determined by φ. Thus the right side of the equation above is the same regardless of which orthonormal basis $e_1, \dots, e_n$ of V is chosen.

EXERCISES 6.B

1 (a) Suppose $\theta \in \mathbf{R}$. Show that $(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)$ and $(\cos\theta, \sin\theta), (\sin\theta, -\cos\theta)$ are orthonormal bases of $\mathbf{R}^2$.

(b) Show that each orthonormal basis of $\mathbf{R}^2$ is of the form given by one of the two possibilities of part (a).

2 Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in V. Let $v \in V$. Prove that
$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2$$
if and only if $v \in \operatorname{span}(e_1, \dots, e_m)$.

3 Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ has an upper-triangular matrix with respect to the basis $(1, 0, 0)$, $(1, 1, 1)$, $(1, 1, 2)$. Find an orthonormal basis of $\mathbf{R}^3$ (use the usual inner product on $\mathbf{R}^3$) with respect to which T has an upper-triangular matrix.

4 Suppose n is a positive integer. Prove that

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\sin nx}{\sqrt{\pi}}$$

is an orthonormal list of vectors in $C[-\pi, \pi]$, the vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

[*The orthonormal list above is often used for modeling periodic phenomena such as tides.*]

5 On $\mathcal{P}_2(\mathbf{R})$, consider the inner product given by

$$\langle p, q \rangle = \int_0^1 p(x)q(x)\,dx.$$

Apply the Gram–Schmidt Procedure to the basis $1, x, x^2$ to produce an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$.

6 Find an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$ (with inner product as in Exercise 5) such that the differentiation operator (the operator that takes p to p') on $\mathcal{P}_2(\mathbf{R})$ has an upper-triangular matrix with respect to this basis.

7 Find a polynomial $q \in \mathcal{P}_2(\mathbf{R})$ such that

$$p\left(\tfrac{1}{2}\right) = \int_0^1 p(x)q(x)\,dx$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

8 Find a polynomial $q \in \mathcal{P}_2(\mathbf{R})$ such that

$$\int_0^1 p(x)(\cos \pi x)\,dx = \int_0^1 p(x)q(x)\,dx$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

9 What happens if the Gram–Schmidt Procedure is applied to a list of vectors that is not linearly independent?

10 Suppose V is a real inner product space and $v_1, \dots, v_m$ is a linearly independent list of vectors in V. Prove that there exist exactly 2^m orthonormal lists $e_1, \dots, e_m$ of vectors in V such that

$$\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$$

for all $j \in \{1, \dots, m\}$.

11 Suppose $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ are inner products on V such that $\langle v, w \rangle_1 = 0$ if and only if $\langle v, w \rangle_2 = 0$. Prove that there is a positive number c such that $\langle v, w \rangle_1 = c \langle v, w \rangle_2$ for every $v, w \in V$.

12 Suppose V is finite-dimensional and $\langle \cdot, \cdot \rangle_1$, $\langle \cdot, \cdot \rangle_2$ are inner products on V with corresponding norms $\| \cdot \|_1$ and $\| \cdot \|_2$. Prove that there exists a positive number c such that

$$\|v\|_1 \leq c \|v\|_2$$

for every $v \in V$.

13 Suppose $v_1, \dots, v_m$ is a linearly independent list in V. Show that there exists $w \in V$ such that $\langle w, v_j \rangle > 0$ for all $j \in \{1, \dots, m\}$.

14 Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $v_1, \dots, v_n$ are vectors in V such that

$$\|e_j - v_j\| < \frac{1}{\sqrt{n}}$$

for each j. Prove that $v_1, \dots, v_n$ is a basis of V.

15 Suppose $C_{\mathbf{R}}([-1, 1])$ is the vector space of continuous real-valued functions on the interval $[-1, 1]$ with inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) \, dx$$

for $f, g \in C_{\mathbf{R}}([-1, 1])$. Let φ be the linear functional on $C_{\mathbf{R}}([-1, 1])$ defined by $\varphi(f) = f(0)$. Show that there does not exist $g \in C_{\mathbf{R}}([-1, 1])$ such that

$$\varphi(f) = \langle f, g \rangle$$

for every $f \in C_{\mathbf{R}}([-1, 1])$.
[*The exercise above shows that the Riesz Representation Theorem* (6.42) *does not hold on infinite-dimensional vector spaces without additional hypotheses on* V *and* φ.]

16 Suppose $\mathbf{F} = \mathbf{C}$, V is finite-dimensional, $T \in \mathcal{L}(V)$, all the eigenvalues of T have absolute value less than 1, and $\epsilon > 0$. Prove that there exists a positive integer m such that $\|T^m v\| \leq \epsilon \|v\|$ for every $v \in V$.

17 For $u \in V$, let Φu denote the linear functional on V defined by

$$(\Phi u)(v) = \langle v, u \rangle$$

for $v \in V$.

(a) Show that if $\mathbf{F} = \mathbf{R}$, then Φ is a linear map from V to V'. (Recall from Section 3.F that $V' = \mathcal{L}(V, \mathbf{F})$ and that V' is called the dual space of V.)

(b) Show that if $\mathbf{F} = \mathbf{C}$ and $V \neq \{0\}$, then Φ is not a linear map.

(c) Show that Φ is injective.

(d) Suppose $\mathbf{F} = \mathbf{R}$ and V is finite-dimensional. Use parts (a) and (c) and a dimension-counting argument (but without using 6.42) to show that Φ is an isomorphism from V onto V'.

[*Part* (*d*) *gives an alternative proof of the Riesz Representation Theorem* (6.42) *when* $\mathbf{F} = \mathbf{R}$. *Part* (*d*) *also gives a natural isomorphism* (*meaning that it does not depend on a choice of basis*) *from a finite-dimensional real inner product space onto its dual space.*]

6.C *Orthogonal Complements and Minimization Problems*

Orthogonal Complements

6.45 Definition ***orthogonal complement,*** $U^\perp$

If U is a subset of V, then the ***orthogonal complement*** of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u\rangle = 0 \text{ for every } u \in U\}.$$

For example, if U is a line in $\mathbf{R}^3$, then $U^\perp$ is the plane containing the origin that is perpendicular to U. If U is a plane in $\mathbf{R}^3$, then $U^\perp$ is the line containing the origin that is perpendicular to U.

6.46 Basic properties of orthogonal complement

(a) If U is a subset of V, then $U^\perp$ is a subspace of V.

(b) $\{0\}^\perp = V$.

(c) $V^\perp = \{0\}$.

(d) If U is a subset of V, then $U \cap U^\perp \subset \{0\}$.

(e) If U and W are subsets of V and $U \subset W$, then $W^\perp \subset U^\perp$.

Proof

(a) Suppose U is a subset of V. Then $\langle 0, u\rangle = 0$ for every $u \in U$; thus $0 \in U^\perp$.

Suppose $v, w \in U^\perp$. If $u \in U$, then

$$\langle v + w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = 0 + 0 = 0.$$

Thus $v + w \in U^\perp$. In other words, $U^\perp$ is closed under addition.

Similarly, suppose $\lambda \in \mathbf{F}$ and $v \in U^\perp$. If $u \in U$, then

$$\langle \lambda v, u\rangle = \lambda\langle v, u\rangle = \lambda \cdot 0 = 0.$$

Thus $\lambda v \in U^\perp$. In other words, $U^\perp$ is closed under scalar multiplication. Thus $U^\perp$ is a subspace of V.

(b) Suppose $v \in V$. Then $\langle v, 0\rangle = 0$, which implies that $v \in \{0\}^\perp$. Thus $\{0\}^\perp = V$.

(c) Suppose $v \in V^\perp$. Then $\langle v, v\rangle = 0$, which implies that $v = 0$. Thus $V^\perp = \{0\}$.

(d) Suppose U is a subset of V and $v \in U \cap U^\perp$. Then $\langle v, v\rangle = 0$, which implies that $v = 0$. Thus $U \cap U^\perp \subset \{0\}$.

(e) Suppose U and W are subsets of V and $U \subset W$. Suppose $v \in W^\perp$. Then $\langle v, u\rangle = 0$ for every $u \in W$, which implies that $\langle v, u\rangle = 0$ for every $u \in U$. Hence $v \in U^\perp$. Thus $W^\perp \subset U^\perp$. ■

Recall that if U, W are subspaces of V, then V is the direct sum of U and W (written $V = U \oplus W$) if each element of V can be written in exactly one way as a vector in U plus a vector in W (see 1.40).

The next result shows that every finite-dimensional subspace of V leads to a natural direct sum decomposition of V.

6.47 Direct sum of a subspace and its orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^\perp.$$

Proof First we will show that

6.48
$$V = U + U^\perp.$$

To do this, suppose $v \in V$. Let $e_1, \dots, e_m$ be an orthonormal basis of U. Obviously

6.49
$$v = \underbrace{\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m}_{u} + \underbrace{v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m}_{w}.$$

Let u and w be defined as in the equation above. Clearly $u \in U$. Because $e_1, \dots, e_m$ is an orthonormal list, for each $j = 1, \dots, m$ we have

$$\begin{aligned}\langle w, e_j\rangle &= \langle v, e_j\rangle - \langle v, e_j\rangle \\ &= 0.\end{aligned}$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \dots, e_m)$. In other words, $w \in U^\perp$. Thus we have written $v = u + w$, where $u \in U$ and $w \in U^\perp$, completing the proof of 6.48.

From 6.46(d), we know that $U \cap U^\perp = \{0\}$. Along with 6.48, this implies that $V = U \oplus U^\perp$ (see 1.45). ■

Now we can see how to compute $\dim U^{\perp}$ from $\dim U$.

6.50 Dimension of the orthogonal complement

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^{\perp} = \dim V - \dim U.$$

Proof The formula for $\dim U^{\perp}$ follows immediately from 6.47 and 3.78. ■

The next result is an important consequence of 6.47.

6.51 The orthogonal complement of the orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$U = (U^{\perp})^{\perp}.$$

Proof First we will show that

6.52 $$U \subset (U^{\perp})^{\perp}.$$

To do this, suppose $u \in U$. Then $\langle u, v\rangle = 0$ for every $v \in U^{\perp}$ (by the definition of $U^{\perp}$). Because u is orthogonal to every vector in $U^{\perp}$, we have $u \in (U^{\perp})^{\perp}$, completing the proof of 6.52.

To prove the inclusion in the other direction, suppose $v \in (U^{\perp})^{\perp}$. By 6.47, we can write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. We have $v - u = w \in U^{\perp}$. Because $v \in (U^{\perp})^{\perp}$ and $u \in (U^{\perp})^{\perp}$ (from 6.52), we have $v - u \in (U^{\perp})^{\perp}$. Thus $v - u \in U^{\perp} \cap (U^{\perp})^{\perp}$, which implies that $v - u$ is orthogonal to itself, which implies that $v - u = 0$, which implies that $v = u$, which implies that $v \in U$. Thus $(U^{\perp})^{\perp} \subset U$, which along with 6.52 completes the proof. ■

We now define an operator P_U for each finite-dimensional subspace of V.

6.53 Definition ***orthogonal projection,*** P_U

Suppose U is a finite-dimensional subspace of V. The ***orthogonal projection*** of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. Then $P_U v = u$.

The direct sum decomposition $V = U \oplus U^\perp$ given by 6.47 shows that each $v \in V$ can be uniquely written in the form $v = u + w$ with $u \in U$ and $w \in U^\perp$. Thus $P_U v$ is well defined.

6.54 **Example** Suppose $x \in V$ with $x \neq 0$ and $U = \operatorname{span}(x)$. Show that

$$P_U v = \frac{\langle v, x \rangle}{\|x\|^2} x$$

for every $v \in V$.

Solution Suppose $v \in V$. Then

$$v = \frac{\langle v, x \rangle}{\|x\|^2} x + \left(v - \frac{\langle v, x \rangle}{\|x\|^2} x \right),$$

where the first term on the right is in $\operatorname{span}(x)$ (and thus in U) and the second term on the right is orthogonal to x (and thus is in $U^\perp$). Thus $P_U v$ equals the first term on the right, as desired.

6.55 Properties of the orthogonal projection P_U

Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

(a) $P_U \in \mathcal{L}(V)$;

(b) $P_U u = u$ for every $u \in U$;

(c) $P_U w = 0$ for every $w \in U^\perp$;

(d) $\operatorname{range} P_U = U$;

(e) $\operatorname{null} P_U = U^\perp$;

(f) $v - P_U v \in U^\perp$;

(g) $P_U{}^2 = P_U$;

(h) $\|P_U v\| \leq \|v\|$;

(i) for every orthonormal basis $e_1, \ldots, e_m$ of U,

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Proof

(a) To show that P_U is a linear map on V, suppose $v_1, v_2 \in V$. Write

$$v_1 = u_1 + w_1 \quad \text{and} \quad v_2 = u_2 + w_2$$

with $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$. Thus $P_U v_1 = u_1$ and $P_U v_2 = u_2$. Now

$$v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2),$$

where $u_1 + u_2 \in U$ and $w_1 + w_2 \in U^\perp$. Thus

$$P_U(v_1 + v_2) = u_1 + u_2 = P_U v_1 + P_U v_2.$$

Similarly, suppose $\lambda \in \mathbf{F}$. The equation $v = u + w$ with $u \in U$ and $w \in U^\perp$ implies that $\lambda v = \lambda u + \lambda w$ with $\lambda u \in U$ and $\lambda w \in U^\perp$. Thus $P_U(\lambda v) = \lambda u = \lambda P_U v$.

Hence P_U is a linear map from V to V.

(b) Suppose $u \in U$. We can write $u = u + 0$, where $u \in U$ and $0 \in U^\perp$. Thus $P_U u = u$.

(c) Suppose $w \in U^\perp$. We can write $w = 0 + w$, where $0 \in U$ and $w \in U^\perp$. Thus $P_U w = 0$.

(d) The definition of P_U implies that range $P_U \subset U$. Part (b) implies that $U \subset$ range P_U. Thus range $P_U = U$.

(e) Part (c) implies that $U^\perp \subset$ null P_U. To prove the inclusion in the other direction, note that if $v \in$ null P_U then the decomposition given by 6.47 must be $v = 0 + v$, where $0 \in U$ and $v \in U^\perp$. Thus null $P_U \subset U^\perp$.

(f) If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$v - P_U v = v - u = w \in U^\perp.$$

(g) If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$({P_U}^2)v = P_U(P_U v) = P_U u = u = P_U v.$$

(h) If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$\|P_U v\|^2 = \|u\|^2 \le \|u\|^2 + \|w\|^2 = \|v\|^2,$$

where the last equality comes from the Pythagorean Theorem.

(i) The formula for $P_U v$ follows from equation 6.49 in the proof of 6.47. ■

Minimization Problems

The remarkable simplicity of the solution to this minimization problem has led to many important applications of inner product spaces outside of pure mathematics.

The following problem often arises: given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. The next proposition shows that this minimization problem is solved by taking $u = P_U v$.

6.56 Minimizing the distance to a subspace

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \leq \|v - u\|.$$

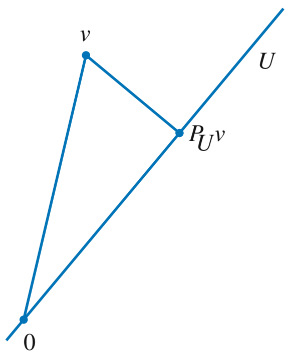
Furthermore, the inequality above is an equality if and only if $u = P_U v$.

Proof We have

6.57
$$\begin{aligned} \|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2, \end{aligned}$$

where the first line above holds because $0 \leq \|P_U v - u\|^2$, the second line above comes from the Pythagorean Theorem [which applies because $v - P_U v \in U^\perp$ by 6.55(f), and $P_U v - u \in U$], and the third line above holds by simple algebra. Taking square roots gives the desired inequality.

Our inequality above is an equality if and only if 6.57 is an equality, which happens if and only if $\|P_U v - u\| = 0$, which happens if and only if $u = P_U v$. ∎



$P_U v$ is the closest point in U to v.

The last result is often combined with the formula 6.55(i) to compute explicit solutions to minimization problems.

6.58 **Example** Find a polynomial u with real coefficients and degree at most 5 that approximates $\sin x$ as well as possible on the interval $[-\pi, \pi]$, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$$

is as small as possible. Compare this result to the Taylor series approximation.

Solution Let $C_{\mathbf{R}}[-\pi, \pi]$ denote the real inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

6.59 $$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\, dx.$$

Let $v \in C_{\mathbf{R}}[-\pi, \pi]$ be the function defined by $v(x) = \sin x$. Let U denote the subspace of $C_{\mathbf{R}}[-\pi, \pi]$ consisting of the polynomials with real coefficients and degree at most 5. Our problem can now be reformulated as follows:

Find $u \in U$ such that $\|v - u\|$ is as small as possible.

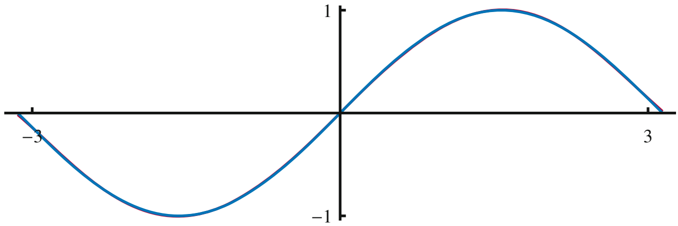
A computer that can perform integrations is useful here.

To compute the solution to our approximation problem, first apply the Gram–Schmidt Procedure (using the inner product given by 6.59) to the basis $1, x, x^2, x^3, x^4, x^5$ of U, producing an orthonormal basis $e_1, e_2, e_3, e_4, e_5, e_6$ of U. Then, again using the inner product given by 6.59, compute $P_U v$ using 6.55(i) (with $m = 6$). Doing this computation shows that $P_U v$ is the function u defined by

6.60 $$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5,$$

where the π's that appear in the exact answer have been replaced with a good decimal approximation.

By 6.56, the polynomial u above is the best approximation to $\sin x$ on $[-\pi, \pi]$ using polynomials of degree at most 5 (here "best approximation" means in the sense of minimizing $\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$). To see how good this approximation is, the next figure shows the graphs of both $\sin x$ and our approximation $u(x)$ given by 6.60 over the interval $[-\pi, \pi]$.



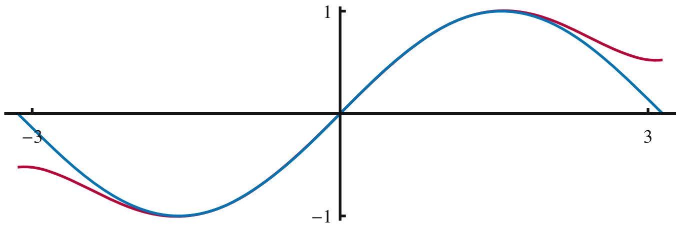
Graphs on $[-\pi, \pi]$ *of* $\sin x$ (*blue*) *and its approximation* $u(x)$ (*red*) *given by 6.60.*

Our approximation 6.60 is so accurate that the two graphs are almost identical—our eyes may see only one graph! Here the blue graph is placed almost exactly over the red graph. If you are viewing this on an electronic device, try enlarging the picture above, especially near 3 or -3, to see a small gap between the two graphs.

Another well-known approximation to $\sin x$ by a polynomial of degree 5 is given by the Taylor polynomial

6.61
$$x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

To see how good this approximation is, the next picture shows the graphs of both $\sin x$ and the Taylor polynomial 6.61 over the interval $[-\pi, \pi]$.



Graphs on $[-\pi, \pi]$ *of* $\sin x$ (*blue*) *and the Taylor polynomial 6.61* (*red*).

The Taylor polynomial is an excellent approximation to $\sin x$ for x near 0. But the picture above shows that for $|x| > 2$, the Taylor polynomial is not so accurate, especially compared to 6.60. For example, taking $x = 3$, our approximation 6.60 estimates $\sin 3$ with an error of about 0.001, but the Taylor series 6.61 estimates $\sin 3$ with an error of about 0.4. Thus at $x = 3$, the error in the Taylor series is hundreds of times larger than the error given by 6.60. Linear algebra has helped us discover an approximation to $\sin x$ that improves upon what we learned in calculus!

EXERCISES 6.C

1 Suppose $v_1, \dots, v_m \in V$. Prove that

$$\{v_1, \dots, v_m\}^{\perp} = \big(\operatorname{span}(v_1, \dots, v_m)\big)^{\perp}.$$

2 Suppose U is a finite-dimensional subspace of V. Prove that $U^{\perp} = \{0\}$ if and only if $U = V$.
[*Exercise 14*(a) *shows that the result above is not true without the hypothesis that U is finite-dimensional.*]

3 Suppose U is a subspace of V with basis $u_1, \dots, u_m$ and

$$u_1, \dots, u_m, w_1, \dots, w_n$$

is a basis of V. Prove that if the Gram–Schmidt Procedure is applied to the basis of V above, producing a list $e_1, \dots, e_m, f_1, \dots, f_n$, then $e_1, \dots, e_m$ is an orthonormal basis of U and $f_1, \dots, f_n$ is an orthonormal basis of $U^{\perp}$.

4 Suppose U is the subspace of $\mathbf{R}^4$ defined by

$$U = \operatorname{span}\big((1, 2, 3, -4), (-5, 4, 3, 2)\big).$$

Find an orthonormal basis of U and an orthonormal basis of $U^{\perp}$.

5 Suppose V is finite-dimensional and U is a subspace of V. Show that $P_{U^{\perp}} = I - P_U$, where I is the identity operator on V.

6 Suppose U and W are finite-dimensional subspaces of V. Prove that $P_U P_W = 0$ if and only if $\langle u, w \rangle = 0$ for all $u \in U$ and all $w \in W$.

7 Suppose V is finite-dimensional and $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and every vector in null P is orthogonal to every vector in range P. Prove that there exists a subspace U of V such that $P = P_U$.

8 Suppose V is finite-dimensional and $P \in \mathcal{L}(V)$ is such that $P^2 = P$ and

$$\|Pv\| \le \|v\|$$

for every $v \in V$. Prove that there exists a subspace U of V such that $P = P_U$.

9 Suppose $T \in \mathcal{L}(V)$ and U is a finite-dimensional subspace of V. Prove that U is invariant under T if and only if $P_U T P_U = T P_U$.

10 Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V. Prove that U and $U^{\perp}$ are both invariant under T if and only if $P_U T = T P_U$.

11 In $\mathbf{R}^4$, let

$$U = \operatorname{span}\bigl((1,1,0,0),(1,1,1,2)\bigr).$$

Find $u \in U$ such that $\|u - (1,2,3,4)\|$ is as small as possible.

12 Find $p \in \mathcal{P}_3(\mathbf{R})$ such that $p(0) = 0$, $p'(0) = 0$, and

$$\int_0^1 |2 + 3x - p(x)|^2 \, dx$$

is as small as possible.

13 Find $p \in \mathcal{P}_5(\mathbf{R})$ that makes

$$\int_{-\pi}^{\pi} |\sin x - p(x)|^2 \, dx$$

as small as possible.
[*The polynomial 6.60 is an excellent approximation to the answer to this exercise, but here you are asked to find the exact solution, which involves powers of π. A computer that can perform symbolic integration will be useful.*]

14 Suppose $C_{\mathbf{R}}([-1,1])$ is the vector space of continuous real-valued functions on the interval $[-1,1]$ with inner product given by

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$$

for $f, g \in C_{\mathbf{R}}([-1,1])$. Let U be the subspace of $C_{\mathbf{R}}([-1,1])$ defined by

$$U = \{f \in C_{\mathbf{R}}([-1,1]) : f(0) = 0\}.$$

(a) Show that $U^{\perp} = \{0\}$.

(b) Show that 6.47 and 6.51 do not hold without the finite-dimensional hypothesis.

CHAPTER

7

Isaac Newton (1642–1727), as envisioned by British poet and artist William Blake in this 1795 painting.

Operators on Inner Product Spaces

The deepest results related to inner product spaces deal with the subject to which we now turn—operators on inner product spaces. By exploiting properties of the adjoint, we will develop a detailed description of several important classes of operators on inner product spaces.

A new assumption for this chapter is listed in the second bullet point below:

7.1 Notation $\mathbf{F}$, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V and W denote finite-dimensional inner product spaces over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- adjoint
- Spectral Theorem
- positive operators
- isometries
- Polar Decomposition
- Singular Value Decomposition

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_7

7.A Self-Adjoint and Normal Operators

Adjoints

7.2 Definition ***adjoint***, T^*

Suppose $T \in \mathcal{L}(V, W)$. The ***adjoint*** of T is the function $T^*: W \to V$ such that

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for every $v \in V$ and every $w \in W$.

*The word **adjoint** has another meaning in linear algebra. We do not need the second meaning in this book. In case you encounter the second meaning for adjoint elsewhere, be warned that the two meanings for adjoint are unrelated to each other.*

To see why the definition above makes sense, suppose $T \in \mathcal{L}(V, W)$. Fix $w \in W$. Consider the linear functional on V that maps $v \in V$ to $\langle Tv, w\rangle$; this linear functional depends on T and w. By the Riesz Representation Theorem (6.42), there exists a unique vector in V such that this linear functional is given by taking the inner product with it. We call this unique vector T^*w. In other words, T^*w is the unique vector in V such that $\langle Tv, w\rangle = \langle v, T^*w\rangle$ for every $v \in V$.

7.3 Example Define $T: \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

Find a formula for T^*.

Solution Here T^* will be a function from $\mathbf{R}^2$ to $\mathbf{R}^3$. To compute T^*, fix a point $(y_1, y_2) \in \mathbf{R}^2$. Then for every $(x_1, x_2, x_3) \in \mathbf{R}^3$ we have

$$\begin{aligned}
\langle (x_1, x_2, x_3), T^*(y_1, y_2)\rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2)\rangle \\
&= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle \\
&= x_2 y_1 + 3x_3 y_1 + 2x_1 y_2 \\
&= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle.
\end{aligned}$$

Thus

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

7.4 Example Fix $u \in V$ and $x \in W$. Define $T \in \mathcal{L}(V, W)$ by

$$Tv = \langle v, u \rangle x$$

for every $v \in V$. Find a formula for T^*.

Solution Fix $w \in W$. Then for every $v \in V$ we have

$$\begin{aligned} \langle v, T^* w \rangle &= \langle Tv, w \rangle \\ &= \big\langle \langle v, u \rangle x, w \big\rangle \\ &= \langle v, u \rangle \langle x, w \rangle \\ &= \big\langle v, \langle w, x \rangle u \big\rangle. \end{aligned}$$

Thus

$$T^* w = \langle w, x \rangle u.$$

In the two examples above, T^* turned out to be not just a function but a linear map. This is true in general, as shown by the next result.

The proofs of the next two results use a common technique: flip T^* from one side of an inner product to become T on the other side.

7.5 The adjoint is a linear map

If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.

Proof Suppose $T \in \mathcal{L}(V, W)$. Fix $w_1, w_2 \in W$. If $v \in V$, then

$$\begin{aligned} \langle v, T^*(w_1 + w_2) \rangle &= \langle Tv, w_1 + w_2 \rangle \\ &= \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle \\ &= \langle v, T^* w_1 \rangle + \langle v, T^* w_2 \rangle \\ &= \langle v, T^* w_1 + T^* w_2 \rangle, \end{aligned}$$

which shows that $T^*(w_1 + w_2) = T^* w_1 + T^* w_2$.

Fix $w \in W$ and $\lambda \in \mathbf{F}$. If $v \in V$, then

$$\begin{aligned} \langle v, T^*(\lambda w) \rangle &= \langle Tv, \lambda w \rangle \\ &= \bar{\lambda} \langle Tv, w \rangle \\ &= \bar{\lambda} \langle v, T^* w \rangle \\ &= \langle v, \lambda T^* w \rangle, \end{aligned}$$

which shows that $T^*(\lambda w) = \lambda T^* w$.

Thus T^* is a linear map, as desired. ■

7.6 Properties of the adjoint

(a) $(S + T)^* = S^* + T^*$ for all $S, T \in \mathcal{L}(V, W)$;

(b) $(\lambda T)^* = \bar{\lambda} T^*$ for all $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$;

(c) $(T^*)^* = T$ for all $T \in \mathcal{L}(V, W)$;

(d) $I^* = I$, where I is the identity operator on V;

(e) $(ST)^* = T^* S^*$ for all $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$ (here U is an inner product space over $\mathbf{F}$).

Proof

(a) Suppose $S, T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\begin{aligned}\langle v, (S + T)^* w\rangle &= \langle (S + T)v, w\rangle \\ &= \langle Sv, w\rangle + \langle Tv, w\rangle \\ &= \langle v, S^* w\rangle + \langle v, T^* w\rangle \\ &= \langle v, S^* w + T^* w\rangle.\end{aligned}$$

Thus $(S + T)^* w = S^* w + T^* w$, as desired.

(b) Suppose $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\langle v, (\lambda T)^* w\rangle = \langle \lambda T v, w\rangle = \lambda \langle Tv, w\rangle = \lambda \langle v, T^* w\rangle = \langle v, \bar{\lambda} T^* w\rangle.$$

Thus $(\lambda T)^* w = \bar{\lambda} T^* w$, as desired.

(c) Suppose $T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\langle w, (T^*)^* v\rangle = \langle T^* w, v\rangle = \overline{\langle v, T^* w\rangle} = \overline{\langle Tv, w\rangle} = \langle w, Tv\rangle.$$

Thus $(T^*)^* v = Tv$, as desired.

(d) If $v, u \in V$, then

$$\langle v, I^* u\rangle = \langle Iv, u\rangle = \langle v, u\rangle.$$

Thus $I^* u = u$, as desired.

(e) Suppose $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$. If $v \in V$ and $u \in U$, then

$$\langle v, (ST)^* u\rangle = \langle STv, u\rangle = \langle Tv, S^* u\rangle = \langle v, T^*(S^* u)\rangle.$$

Thus $(ST)^* u = T^*(S^* u)$, as desired. ■

The next result shows the relationship between the null space and the range of a linear map and its adjoint. The symbol $\Longleftrightarrow$ used in the proof means "if and only if"; this symbol could also be read to mean "is equivalent to".

7.7 Null space and range of T^*

Suppose $T \in \mathcal{L}(V, W)$. Then

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Proof We begin by proving (a). Let $w \in W$. Then

$$\begin{aligned} w \in \operatorname{null} T^* &\Longleftrightarrow T^* w = 0 \\ &\Longleftrightarrow \langle v, T^* w \rangle = 0 \text{ for all } v \in V \\ &\Longleftrightarrow \langle T v, w \rangle = 0 \text{ for all } v \in V \\ &\Longleftrightarrow w \in (\operatorname{range} T)^\perp. \end{aligned}$$

Thus $\operatorname{null} T^* = (\operatorname{range} T)^\perp$, proving (a).

If we take the orthogonal complement of both sides of (a), we get (d), where we have used 6.51. Replacing T with T^* in (a) gives (c), where we have used 7.6(c). Finally, replacing T with T^* in (d) gives (b). ∎

7.8 Definition ***conjugate transpose***

The ***conjugate transpose*** of an m-by-n matrix is the n-by-m matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry.

7.9 Example

The conjugate transpose of the matrix $\begin{pmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{pmatrix}$ is the matrix

$$\begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}.$$

*If $\mathbf{F} = \mathbf{R}$, then the conjugate transpose of a matrix is the same as its **transpose**, which is the matrix obtained by interchanging the rows and columns.*

The adjoint of a linear map does not depend on a choice of basis. This explains why this book emphasizes adjoints of linear maps instead of conjugate transposes of matrices.

The next result shows how to compute the matrix of T^* from the matrix of T.

Caution: Remember that the result below applies only when we are dealing with orthonormal bases. With respect to nonorthonormal bases, the matrix of T^* does not necessarily equal the conjugate transpose of the matrix of T.

7.10 The matrix of T^*

Let $T \in \mathcal{L}(V, W)$. Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $f_1, \ldots, f_m$ is an orthonormal basis of W. Then

$$\mathcal{M}\big(T^*, (f_1, \ldots, f_m), (e_1, \ldots, e_n)\big)$$

is the conjugate transpose of

$$\mathcal{M}\big(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m)\big).$$

Proof In this proof, we will write $\mathcal{M}(T)$ instead of the longer expression $\mathcal{M}\big(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m)\big)$; we will also write $\mathcal{M}(T^*)$ instead of $\mathcal{M}\big(T^*, (f_1, \ldots, f_m), (e_1, \ldots, e_n)\big)$.

Recall that we obtain the k^{th} column of $\mathcal{M}(T)$ by writing Te_k as a linear combination of the f_j's; the scalars used in this linear combination then become the k^{th} column of $\mathcal{M}(T)$. Because $f_1, \ldots, f_m$ is an orthonormal basis of W, we know how to write Te_k as a linear combination of the f_j's (see 6.30):

$$Te_k = \langle Te_k, f_1 \rangle f_1 + \cdots + \langle Te_k, f_m \rangle f_m.$$

Thus the entry in row j, column k, of $\mathcal{M}(T)$ is $\langle Te_k, f_j \rangle$.

Replacing T with T^* and interchanging the roles played by the e's and f's, we see that the entry in row j, column k, of $\mathcal{M}(T^*)$ is $\langle T^* f_k, e_j \rangle$, which equals $\langle f_k, Te_j \rangle$, which equals $\overline{\langle Te_j, f_k \rangle}$, which equals the complex conjugate of the entry in row k, column j, of $\mathcal{M}(T)$. In other words, $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$. ■

Self-Adjoint Operators

Now we switch our attention to operators on inner product spaces. Thus instead of considering linear maps from V to W, we will be focusing on linear maps from V to V; recall that such linear maps are called operators.

7.11 Definition ***self-adjoint***

An operator $T \in \mathcal{L}(V)$ is called ***self-adjoint*** if $T = T^*$. In other words, $T \in \mathcal{L}(V)$ is self-adjoint if and only if

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

for all $v, w \in V$.

7.12 Example Suppose T is the operator on $\mathbf{F}^2$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & b \\ 3 & 7 \end{pmatrix}.$$

Find all numbers b such that T is self-adjoint.

Solution The operator T is self-adjoint if and only if $b = 3$ (because $\mathcal{M}(T) = \mathcal{M}(T^*)$ if and only if $b = 3$; recall that $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$—see 7.10).

You should verify that the sum of two self-adjoint operators is self-adjoint and that the product of a real scalar and a self-adjoint operator is self-adjoint.

A good analogy to keep in mind (especially when $\mathbf{F} = \mathbf{C}$) is that the adjoint on $\mathcal{L}(V)$ plays a role similar to complex conjugation on $\mathbf{C}$. A complex number z is real if and only if $z = \bar{z}$; thus a self-adjoint operator ($T = T^*$) is analogous to a real number.

Some mathematicians use the term ***Hermitian*** *instead of self-adjoint, honoring French mathematician Charles Hermite, who in 1873 published the first proof that e is not a zero of any polynomial with integer coefficients.*

We will see that the analogy discussed above is reflected in some important properties of self-adjoint operators, beginning with eigenvalues in the next result.

If $\mathbf{F} = \mathbf{R}$, then by definition every eigenvalue is real, so the next result is interesting only when $\mathbf{F} = \mathbf{C}$.

7.13 Eigenvalues of self-adjoint operators are real

Every eigenvalue of a self-adjoint operator is real.

Proof Suppose T is a self-adjoint operator on V. Let λ be an eigenvalue of T, and let v be a nonzero vector in V such that $Tv = \lambda v$. Then

$$\lambda\|v\|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \bar{\lambda}\|v\|^2.$$

Thus $\lambda = \bar{\lambda}$, which means that λ is real, as desired. ■

The next result is false for real inner product spaces. As an example, consider the operator $T \in \mathcal{L}(\mathbf{R}^2)$ that is a counterclockwise rotation of 90° around the origin; thus $T(x, y) = (-y, x)$. Obviously Tv is orthogonal to v for every $v \in \mathbf{R}^2$, even though $T \neq 0$.

7.14 Over $\mathbf{C}$, Tv is orthogonal to v for all v only for the 0 operator

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Suppose

$$\langle Tv, v\rangle = 0$$

for all $v \in V$. Then $T = 0$.

Proof We have

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} + \frac{\langle T(u+iw), u+iw\rangle - \langle T(u-iw), u-iw\rangle}{4}\, i$$

for all $u, w \in V$, as can be verified by computing the right side. Note that each term on the right side is of the form $\langle Tv, v\rangle$ for appropriate $v \in V$. Thus our hypothesis implies that $\langle Tu, w\rangle = 0$ for all $u, w \in V$. This implies that $T = 0$ (take $w = Tu$). ■

The next result provides another example of how self-adjoint operators behave like real numbers.

The next result is false for real inner product spaces, as shown by considering any operator on a real inner product space that is not self-adjoint.

7.15 Over $\mathbf{C}$, $\langle Tv, v\rangle$ is real for all v only for self-adjoint operators

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then T is self-adjoint if and only if

$$\langle Tv, v\rangle \in \mathbf{R}$$

for every $v \in V$.

Proof Let $v \in V$. Then

$$\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v\rangle - \langle T^*v, v\rangle = \langle (T - T^*)v, v\rangle.$$

If $\langle Tv, v\rangle \in \mathbf{R}$ for every $v \in V$, then the left side of the equation above equals 0, so $\langle (T - T^*)v, v\rangle = 0$ for every $v \in V$. This implies that $T - T^* = 0$ (by 7.14). Hence T is self-adjoint.

Conversely, if T is self-adjoint, then the right side of the equation above equals 0, so $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle}$ for every $v \in V$. This implies that $\langle Tv, v\rangle \in \mathbf{R}$ for every $v \in V$, as desired. ■

On a real inner product space V, a nonzero operator T might satisfy $\langle Tv, v\rangle = 0$ for all $v \in V$. However, the next result shows that this cannot happen for a self-adjoint operator.

7.16 If $T = T^*$ and $\langle Tv, v\rangle = 0$ for all v, then $T = 0$

Suppose T is a self-adjoint operator on V such that

$$\langle Tv, v\rangle = 0$$

for all $v \in V$. Then $T = 0$.

Proof We have already proved this (without the hypothesis that T is self-adjoint) when V is a complex inner product space (see 7.14). Thus we can assume that V is a real inner product space. If $u, w \in V$, then

7.17 $$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4};$$

this is proved by computing the right side using the equation

$$\langle Tw, u\rangle = \langle w, Tu\rangle = \langle Tu, w\rangle,$$

where the first equality holds because T is self-adjoint and the second equality holds because we are working in a real inner product space.

Each term on the right side of 7.17 is of the form $\langle Tv, v\rangle$ for appropriate v. Thus $\langle Tu, w\rangle = 0$ for all $u, w \in V$. This implies that $T = 0$ (take $w = Tu$). ■

Normal Operators

7.18 Definition ***normal***

- An operator on an inner product space is called ***normal*** if it commutes with its adjoint.
- In other words, $T \in \mathcal{L}(V)$ is normal if

$$TT^* = T^*T.$$

Obviously every self-adjoint operator is normal, because if T is self-adjoint then $T^* = T$.

7.19 Example Let T be the operator on $\mathbf{F}^2$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

Show that T is not self-adjoint and that T is normal.

Solution This operator is not self-adjoint because the entry in row 2, column 1 (which equals 3) does not equal the complex conjugate of the entry in row 1, column 2 (which equals -3).

The matrix of TT^* equals

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}, \text{ which equals } \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Similarly, the matrix of T^*T equals

$$\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}, \text{ which equals } \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Because TT^* and T^*T have the same matrix, we see that $TT^* = T^*T$. Thus T is normal.

The next result implies that null T = null T^* *for every normal operator* T.

In the next section we will see why normal operators are worthy of special attention.

The next result provides a simple characterization of normal operators.

7.20 T is normal if and only if $\|Tv\| = \|T^*v\|$ for all v

An operator $T \in \mathcal{L}(V)$ is normal if and only if

$$\|Tv\| = \|T^*v\|$$

for all $v \in V$.

Proof Let $T \in \mathcal{L}(V)$. We will prove both directions of this result at the same time. Note that

$$\begin{aligned} T \text{ is normal} &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0 \quad \text{for all } v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \quad \text{for all } v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \quad \text{for all } v \in V, \end{aligned}$$

where we used 7.16 to establish the second equivalence (note that the operator $T^*T - TT^*$ is self-adjoint). The equivalence of the first and last conditions above gives the desired result. ■

Compare the next corollary to Exercise 2. That exercise states that the eigenvalues of the adjoint of each operator are equal (as a set) to the complex conjugates of the eigenvalues of the operator. The exercise says nothing about eigenvectors, because an operator and its adjoint may have different eigenvectors. However, the next corollary implies that a normal operator and its adjoint have the same eigenvectors.

7.21 For T normal, T and T^* have the same eigenvectors

Suppose $T \in \mathcal{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T^* with eigenvalue $\bar{\lambda}$.

Proof Because T is normal, so is $T - \lambda I$, as you should verify. Using 7.20, we have

$$0 = \|(T - \lambda I)v\| = \|(T - \lambda I)^*v\| = \|(T^* - \bar{\lambda} I)v\|.$$

Hence v is an eigenvector of T^* with eigenvalue $\bar{\lambda}$, as desired. ■

Because every self-adjoint operator is normal, the next result applies in particular to self-adjoint operators.

7.22 Orthogonal eigenvectors for normal operators

Suppose $T \in \mathcal{L}(V)$ is normal. Then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof Suppose α, β are distinct eigenvalues of T, with corresponding eigenvectors u, v. Thus $Tu = \alpha u$ and $Tv = \beta v$. From 7.21 we have $T^*v = \bar{\beta}v$. Thus

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \bar{\beta}v\rangle \\ &= \langle Tu, v\rangle - \langle u, T^*v\rangle \\ &= 0.\end{aligned}$$

Because $\alpha \neq \beta$, the equation above implies that $\langle u, v\rangle = 0$. Thus u and v are orthogonal, as desired. ■

EXERCISES 7.A

1 Suppose n is a positive integer. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by

$$T(z_1, \dots, z_n) = (0, z_1, \dots, z_{n-1}).$$

Find a formula for $T^*(z_1, \dots, z_n)$.

2 Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Prove that λ is an eigenvalue of T if and only if $\bar{\lambda}$ is an eigenvalue of T^*.

3 Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that U is invariant under T if and only if $U^\perp$ is invariant under T^*.

4 Suppose $T \in \mathcal{L}(V, W)$. Prove that

(a) T is injective if and only if T^* is surjective;

(b) T is surjective if and only if T^* is injective.

5 Prove that

$$\dim \operatorname{null} T^* = \dim \operatorname{null} T + \dim W - \dim V$$

and

$$\dim \operatorname{range} T^* = \dim \operatorname{range} T$$

for every $T \in \mathcal{L}(V, W)$.

6 Make $\mathcal{P}_2(\mathbf{R})$ into an inner product space by defining

$$\langle p, q \rangle = \int_0^1 p(x)q(x)\,dx.$$

Define $T \in \mathcal{L}\big(\mathcal{P}_2(\mathbf{R})\big)$ by $T(a_0 + a_1 x + a_2 x^2) = a_1 x$.

(a) Show that T is not self-adjoint.

(b) The matrix of T with respect to the basis $(1, x, x^2)$ is

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This matrix equals its conjugate transpose, even though T is not self-adjoint. Explain why this is not a contradiction.

7 Suppose $S, T \in \mathcal{L}(V)$ are self-adjoint. Prove that ST is self-adjoint if and only if $ST = TS$.

8 Suppose V is a real inner product space. Show that the set of self-adjoint operators on V is a subspace of $\mathcal{L}(V)$.

9 Suppose V is a complex inner product space with $V \neq \{0\}$. Show that the set of self-adjoint operators on V is not a subspace of $\mathcal{L}(V)$.

10 Suppose $\dim V \geq 2$. Show that the set of normal operators on V is not a subspace of $\mathcal{L}(V)$.

11 Suppose $P \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that there is a subspace U of V such that $P = P_U$ if and only if P is self-adjoint.

12 Suppose that T is a normal operator on V and that 3 and 4 are eigenvalues of T. Prove that there exists a vector $v \in V$ such that $\|v\| = \sqrt{2}$ and $\|Tv\| = 5$.

13 Give an example of an operator $T \in \mathcal{L}(\mathbf{C}^4)$ such that T is normal but not self-adjoint.

14 Suppose T is a normal operator on V. Suppose also that $v, w \in V$ satisfy the equations

$$\|v\| = \|w\| = 2, \quad Tv = 3v, \quad Tw = 4w.$$

Show that $\|T(v + w)\| = 10$.

15 Fix $u, x \in V$. Define $T \in \mathcal{L}(V)$ by

$$Tv = \langle v, u \rangle x$$

for every $v \in V$.

(a) Suppose $\mathbf{F} = \mathbf{R}$. Prove that T is self-adjoint if and only if u, x is linearly dependent.

(b) Prove that T is normal if and only if u, x is linearly dependent.

16 Suppose $T \in \mathcal{L}(V)$ is normal. Prove that

$$\operatorname{range} T = \operatorname{range} T^*.$$

17 Suppose $T \in \mathcal{L}(V)$ is normal. Prove that

$$\operatorname{null} T^k = \operatorname{null} T \quad \text{and} \quad \operatorname{range} T^k = \operatorname{range} T$$

for every positive integer k.

18 Prove or give a counterexample: If $T \in \mathcal{L}(V)$ and there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Te_j\| = \|T^*e_j\|$ for each j, then T is normal.

19 Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is normal and $T(1,1,1) = (2,2,2)$. Suppose $(z_1, z_2, z_3) \in \operatorname{null} T$. Prove that $z_1 + z_2 + z_3 = 0$.

20 Suppose $T \in \mathcal{L}(V, W)$ and $\mathbf{F} = \mathbf{R}$. Let Φ_V be the isomorphism from V onto the dual space V' given by Exercise 17 in Section 6.B, and let Φ_W be the corresponding isomorphism from W onto W'. Show that if Φ_V and Φ_W are used to identify V and W with V' and W', then T^* is identified with the dual map T'. More precisely, show that $\Phi_V \circ T^* = T' \circ \Phi_W$.

21 Fix a positive integer n. In the inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx,$$

let

$$V = \operatorname{span}(1, \cos x, \cos 2x, \dots, \cos nx, \sin x, \sin 2x, \dots, \sin nx).$$

(a) Define $D \in \mathcal{L}(V)$ by $Df = f'$. Show that $D^* = -D$. Conclude that D is normal but not self-adjoint.

(b) Define $T \in \mathcal{L}(V)$ by $Tf = f''$. Show that T is self-adjoint.

7.B *The Spectral Theorem*

Recall that a diagonal matrix is a square matrix that is 0 everywhere except possibly along the diagonal. Recall also that an operator on V has a diagonal matrix with respect to a basis if and only if the basis consists of eigenvectors of the operator (see 5.41).

The nicest operators on V are those for which there is an *orthonormal* basis of V with respect to which the operator has a diagonal matrix. These are precisely the operators $T \in \mathcal{L}(V)$ such that there is an orthonormal basis of V consisting of eigenvectors of T. Our goal in this section is to prove the Spectral Theorem, which characterizes these operators as the normal operators when $\mathbf{F} = \mathbf{C}$ and as the self-adjoint operators when $\mathbf{F} = \mathbf{R}$. The Spectral Theorem is probably the most useful tool in the study of operators on inner product spaces.

Because the conclusion of the Spectral Theorem depends on $\mathbf{F}$, we will break the Spectral Theorem into two pieces, called the Complex Spectral Theorem and the Real Spectral Theorem. As is often the case in linear algebra, complex vector spaces are easier to deal with than real vector spaces. Thus we present the Complex Spectral Theorem first.

The Complex Spectral Theorem

The key part of the Complex Spectral Theorem (7.24) states that if $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal, then T has a diagonal matrix with respect to some orthonormal basis of V. The next example illustrates this conclusion.

7.23 **Example** Consider the normal operator $T \in \mathcal{L}(\mathbf{C}^2)$ from Example 7.19, whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

As you can verify, $\frac{(i,1)}{\sqrt{2}}, \frac{(-i,1)}{\sqrt{2}}$ is an orthonormal basis of $\mathbf{C}^2$ consisting of eigenvectors of T, and with respect to this basis the matrix of T is the diagonal matrix

$$\begin{pmatrix} 2+3i & 0 \\ 0 & 2-3i \end{pmatrix}.$$

In the next result, the equivalence of (b) and (c) is easy (see 5.41). Thus we prove only that (c) implies (a) and that (a) implies (c).

7.24 Complex Spectral Theorem

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

Proof First suppose (c) holds, so T has a diagonal matrix with respect to some orthonormal basis of V. The matrix of T^* (with respect to the same basis) is obtained by taking the conjugate transpose of the matrix of T; hence T^* also has a diagonal matrix. Any two diagonal matrices commute; thus T commutes with T^*, which means that T is normal. In other words, (a) holds.

Now suppose (a) holds, so T is normal. By Schur's Theorem (6.38), there is an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which T has an upper-triangular matrix. Thus we can write

7.25
$$\mathcal{M}\big(T, (e_1, \ldots, e_n)\big) = \begin{pmatrix} a_{1,1} & \ldots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}.$$

We will show that this matrix is actually a diagonal matrix.

We see from the matrix above that

$$\|Te_1\|^2 = |a_{1,1}|^2$$

and

$$\|T^*e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2.$$

Because T is normal, $\|Te_1\| = \|T^*e_1\|$ (see 7.20). Thus the two equations above imply that all entries in the first row of the matrix in 7.25, except possibly the first entry $a_{1,1}$, equal 0.

Now from 7.25 we see that

$$\|Te_2\|^2 = |a_{2,2}|^2$$

(because $a_{1,2} = 0$, as we showed in the paragraph above) and

$$\|T^*e_2\|^2 = |a_{2,2}|^2 + |a_{2,3}|^2 + \cdots + |a_{2,n}|^2.$$

Because T is normal, $\|Te_2\| = \|T^*e_2\|$. Thus the two equations above imply that all entries in the second row of the matrix in 7.25, except possibly the diagonal entry $a_{2,2}$, equal 0.

Continuing in this fashion, we see that all the nondiagonal entries in the matrix 7.25 equal 0. Thus (c) holds. ■

The Real Spectral Theorem

We will need a few preliminary results, which apply to both real and complex inner product spaces, for our proof of the Real Spectral Theorem.

This technique of completing the square can be used to derive the quadratic formula.

You could guess that the next result is true and even discover its proof by thinking about quadratic polynomials with real coefficients. Specifically, suppose $b, c \in \mathbf{R}$ and $b^2 < 4c$. Let x be a real number. Then

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0.$$

In particular, $x^2 + bx + c$ is an invertible real number (a convoluted way of saying that it is not 0). Replacing the real number x with a self-adjoint operator (recall the analogy between real numbers and self-adjoint operators), we are led to the result below.

7.26 Invertible quadratic expressions

Suppose $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbf{R}$ are such that $b^2 < 4c$. Then

$$T^2 + bT + cI$$

is invertible.

Proof Let v be a nonzero vector in V. Then

$$\begin{aligned}
\langle (T^2 + bT + cI)v, v \rangle &= \langle T^2 v, v \rangle + b\langle Tv, v \rangle + c\langle v, v \rangle \\
&= \langle Tv, Tv \rangle + b\langle Tv, v \rangle + c\|v\|^2 \\
&\geq \|Tv\|^2 - |b|\|Tv\|\|v\| + c\|v\|^2 \\
&= \left(\|Tv\| - \frac{|b|\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\
&> 0,
\end{aligned}$$

where the third line above holds by the Cauchy–Schwarz Inequality (6.15). The last inequality implies that $(T^2 + bT + cI)v \neq 0$. Thus $T^2 + bT + cI$ is injective, which implies that it is invertible (see 3.69). ■

We know that every operator, self-adjoint or not, on a finite-dimensional nonzero complex vector space has an eigenvalue (see 5.21). Thus the next result tells us something new only for real inner product spaces.

7.27 Self-adjoint operators have eigenvalues

Suppose $V \neq \{0\}$ and $T \in \mathcal{L}(V)$ is a self-adjoint operator. Then T has an eigenvalue.

Proof We can assume that V is a real inner product space, as we have already noted. Let $n = \dim V$ and choose $v \in V$ with $v \neq 0$. Then

$$v, Tv, T^2 v, \dots, T^n v$$

cannot be linearly independent, because V has dimension n and we have $n+1$ vectors. Thus there exist real numbers $a_0, \dots, a_n$, not all 0, such that

$$0 = a_0 v + a_1 T v + \dots + a_n T^n v.$$

Make the a's the coefficients of a polynomial, which can be written in factored form (see 4.17) as

$$\begin{aligned} &a_0 + a_1 x + \dots + a_n x^n \\ &\quad = c(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M)(x - \lambda_1) \cdots (x - \lambda_m), \end{aligned}$$

where c is a nonzero real number, each b_j, c_j, and λ_j is real, each $b_j{}^2$ is less than $4c_j$, $m + M \geq 1$, and the equation holds for all real x. We then have

$$\begin{aligned} 0 &= a_0 v + a_1 T v + \dots + a_n T^n v \\ &= (a_0 I + a_1 T + \dots + a_n T^n) v \\ &= c(T^2 + b_1 T + c_1 I) \cdots (T^2 + b_M T + c_M I)(T - \lambda_1 I) \cdots (T - \lambda_m I) v. \end{aligned}$$

By 7.26, each $T^2 + b_j T + c_j I$ is invertible. Recall also that $c \neq 0$. Thus the equation above implies that $m > 0$ and

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_m I) v.$$

Hence $T - \lambda_j I$ is not injective for at least one j. In other words, T has an eigenvalue. ∎

The next result shows that if U is a subspace of V that is invariant under a self-adjoint operator T, then $U^\perp$ is also invariant under T. Later we will show that the hypothesis that T is self-adjoint can be replaced with the weaker hypothesis that T is normal (see 9.30).

7.28 Self-adjoint operators and invariant subspaces

Suppose $T \in \mathcal{L}(V)$ is self-adjoint and U is a subspace of V that is invariant under T. Then

(a) $U^{\perp}$ is invariant under T;

(b) $T|_U \in \mathcal{L}(U)$ is self-adjoint;

(c) $T|_{U^{\perp}} \in \mathcal{L}(U^{\perp})$ is self-adjoint.

Proof To prove (a), suppose $v \in U^{\perp}$. Let $u \in U$. Then

$$\langle Tv, u \rangle = \langle v, Tu \rangle = 0,$$

where the first equality above holds because T is self-adjoint and the second equality above holds because U is invariant under T (and hence $Tu \in U$) and because $v \in U^{\perp}$. Because the equation above holds for each $u \in U$, we conclude that $Tv \in U^{\perp}$. Thus $U^{\perp}$ is invariant under T, completing the proof of (a).

To prove (b), note that if $u, v \in U$, then

$$\langle (T|_U)u, v \rangle = \langle Tu, v \rangle = \langle u, Tv \rangle = \langle u, (T|_U)v \rangle.$$

Thus $T|_U$ is self-adjoint.

Now (c) follows from replacing U with $U^{\perp}$ in (b), which makes sense by (a). ■

We can now prove the next result, which is one of the major theorems in linear algebra.

7.29 Real Spectral Theorem

Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is self-adjoint.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

Proof First suppose (c) holds, so T has a diagonal matrix with respect to some orthonormal basis of V. A diagonal matrix equals its transpose. Hence $T = T^*$, and thus T is self-adjoint. In other words, (a) holds.

We will prove that (a) implies (b) by induction on $\dim V$. To get started, note that if $\dim V = 1$, then (a) implies (b). Now assume that $\dim V > 1$ and that (a) implies (b) for all real inner product spaces of smaller dimension.

Suppose (a) holds, so $T \in \mathcal{L}(V)$ is self-adjoint. Let u be an eigenvector of T with $\|u\| = 1$ (7.27 guarantees that T has an eigenvector, which can then be divided by its norm to produce an eigenvector with norm 1). Let $U = \operatorname{span}(u)$. Then U is a 1-dimensional subspace of V that is invariant under T. By 7.28(c), the operator $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.

By our induction hypothesis, there is an orthonormal basis of $U^\perp$ consisting of eigenvectors of $T|_{U^\perp}$. Adjoining u to this orthonormal basis of $U^\perp$ gives an orthonormal basis of V consisting of eigenvectors of T, completing the proof that (a) implies (b).

We have proved that (c) implies (a) and that (a) implies (b). Clearly (b) implies (c), completing the proof. ■

7.30 **Example** Consider the self-adjoint operator T on $\mathbf{R}^3$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{pmatrix}.$$

As you can verify,

$$\frac{(1,-1,0)}{\sqrt{2}}, \frac{(1,1,1)}{\sqrt{3}}, \frac{(1,1,-2)}{\sqrt{6}}$$

is an orthonormal basis of $\mathbf{R}^3$ consisting of eigenvectors of T, and with respect to this basis, the matrix of T is the diagonal matrix

$$\begin{pmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -15 \end{pmatrix}.$$

If $\mathbf{F} = \mathbf{C}$, then the Complex Spectral Theorem gives a complete description of the normal operators on V. A complete description of the self-adjoint operators on V then easily follows (they are the normal operators on V whose eigenvalues all are real; see Exercise 6).

If $\mathbf{F} = \mathbf{R}$, then the Real Spectral Theorem gives a complete description of the self-adjoint operators on V. In Chapter 9, we will give a complete description of the normal operators on V (see 9.34).

EXERCISES 7.B

1 True or false (and give a proof of your answer): There exists $T \in \mathcal{L}(\mathbf{R}^3)$ such that T is not self-adjoint (with respect to the usual inner product) and such that there is a basis of $\mathbf{R}^3$ consisting of eigenvectors of T.

2 Suppose that T is a self-adjoint operator on a finite-dimensional inner product space and that 2 and 3 are the only eigenvalues of T. Prove that $T^2 - 5T + 6I = 0$.

3 Give an example of an operator $T \in \mathcal{L}(\mathbf{C}^3)$ such that 2 and 3 are the only eigenvalues of T and $T^2 - 5T + 6I \neq 0$.

4 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is normal if and only if all pairs of eigenvectors corresponding to distinct eigenvalues of T are orthogonal and

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

5 Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$. Prove that T is self-adjoint if and only if all pairs of eigenvectors corresponding to distinct eigenvalues of T are orthogonal and

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

6 Prove that a normal operator on a complex inner product space is self-adjoint if and only if all its eigenvalues are real.
[*The exercise above strengthens the analogy (for normal operators) between self-adjoint operators and real numbers.*]

7 Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$ is a normal operator such that $T^9 = T^8$. Prove that T is self-adjoint and $T^2 = T$.

8 Give an example of an operator T on a complex vector space such that $T^9 = T^8$ but $T^2 \neq T$.

9 Suppose V is a complex inner product space. Prove that every normal operator on V has a square root. (An operator $S \in \mathcal{L}(V)$ is called a ***square root*** of $T \in \mathcal{L}(V)$ if $S^2 = T$.)

10 Give an example of a real inner product space V and $T \in \mathcal{L}(V)$ and real numbers b, c with $b^2 < 4c$ such that $T^2 + bT + cI$ is not invertible.
[*The exercise above shows that the hypothesis that T is self-adjoint is needed in 7.26, even for real vector spaces.*]

11 Prove or give a counterexample: every self-adjoint operator on V has a cube root. (An operator $S \in \mathcal{L}(V)$ is called a ***cube root*** of $T \in \mathcal{L}(V)$ if $S^3 = T$.)

12 Suppose $T \in \mathcal{L}(V)$ is self-adjoint, $\lambda \in \mathbf{F}$, and $\epsilon > 0$. Suppose there exists $v \in V$ such that $\|v\| = 1$ and

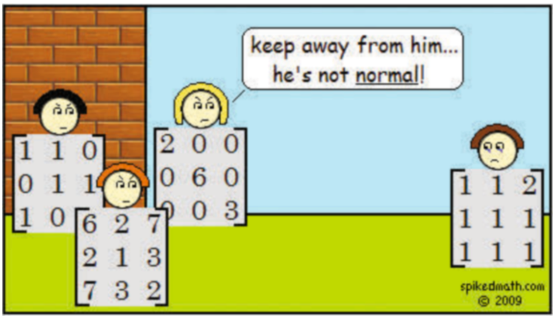
$$\|Tv - \lambda v\| < \epsilon.$$

Prove that T has an eigenvalue λ' such that $|\lambda - \lambda'| < \epsilon$.

13 Give an alternative proof of the Complex Spectral Theorem that avoids Schur's Theorem and instead follows the pattern of the proof of the Real Spectral Theorem.

14 Suppose U is a finite-dimensional real vector space and $T \in \mathcal{L}(U)$. Prove that U has a basis consisting of eigenvectors of T if and only if there is an inner product on U that makes T into a self-adjoint operator.

15 Find the matrix entry below that is covered up.



7.C *Positive Operators and Isometries*

Positive Operators

7.31 Definition ***positive operator***

An operator $T \in \mathcal{L}(V)$ is called ***positive*** if T is self-adjoint and

$$\langle Tv, v \rangle \geq 0$$

for all $v \in V$.

If V is a complex vector space, then the requirement that T is self-adjoint can be dropped from the definition above (by 7.15).

7.32 Example ***positive operators***

(a) If U is a subspace of V, then the orthogonal projection P_U is a positive operator, as you should verify.

(b) If $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbf{R}$ are such that $b^2 < 4c$, then $T^2 + bT + cI$ is a positive operator, as shown by the proof of 7.26.

7.33 Definition ***square root***

An operator R is called a ***square root*** of an operator T if $R^2 = T$.

7.34 Example If $T \in \mathcal{L}(\mathbf{F}^3)$ is defined by $T(z_1, z_2, z_3) = (z_3, 0, 0)$, then the operator $R \in \mathcal{L}(\mathbf{F}^3)$ defined by $R(z_1, z_2, z_3) = (z_2, z_3, 0)$ is a square root of T.

The positive operators correspond to the numbers $[0, \infty)$, so better terminology would use the term nonnegative instead of positive. However, operator theorists consistently call these the positive operators, so we will follow that custom.

The characterizations of the positive operators in the next result correspond to characterizations of the nonnegative numbers among $\mathbf{C}$. Specifically, a complex number z is nonnegative if and only if it has a nonnegative square root, corresponding to condition (c). Also, z is nonnegative if and only if it has a real square root, corresponding to condition (d). Finally, z is nonnegative if and only if there exists a complex number w such that $z = \bar{w}w$, corresponding to condition (e).

7.35 Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all the eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists an operator $R \in \mathcal{L}(V)$ such that $T = R^*R$.

Proof We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (a).

First suppose (a) holds, so that T is positive. Obviously T is self-adjoint (by the definition of a positive operator). To prove the other condition in (b), suppose λ is an eigenvalue of T. Let v be an eigenvector of T corresponding to λ. Then

$$0 \leq \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle.$$

Thus λ is a nonnegative number. Hence (b) holds.

Now suppose (b) holds, so that T is self-adjoint and all the eigenvalues of T are nonnegative. By the Spectral Theorem (7.24 and 7.29), there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of T corresponding to $e_1, \dots, e_n$,; thus each λ_j is a nonnegative number. Let R be the linear map from V to V such that

$$Re_j = \sqrt{\lambda_j}e_j$$

for $j = 1, \dots, n$ (see 3.5). Then R is a positive operator, as you should verify. Furthermore, $R^2e_j = \lambda_j e_j = Te_j$ for each j, which implies that $R^2 = T$. Thus R is a positive square root of T. Hence (c) holds.

Clearly (c) implies (d) (because, by definition, every positive operator is self-adjoint).

Now suppose (d) holds, meaning that there exists a self-adjoint operator R on V such that $T = R^2$. Then $T = R^*R$ (because $R^* = R$). Hence (e) holds.

Finally, suppose (e) holds. Let $R \in \mathcal{L}(V)$ be such that $T = R^*R$. Then $T^* = (R^*R)^* = R^*(R^*)^* = R^*R = T$. Hence T is self-adjoint. To complete the proof that (a) holds, note that

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle \geq 0$$

for every $v \in V$. Thus T is positive. ■

Each nonnegative number has a unique nonnegative square root. The next result shows that positive operators enjoy a similar property.

Some mathematicians also use the term ***positive semidefinite operator****, which means the same as positive operator.*

7.36 Each positive operator has only one positive square root

Every positive operator on V has a unique positive square root.

Proof Suppose $T \in \mathcal{L}(V)$ is positive. Suppose $v \in V$ is an eigenvector of T. Thus there exists $\lambda \geq 0$ such that $Tv = \lambda v$.

A positive operator can have infinitely many square roots (although only one of them can be positive). For example, the identity operator on V has infinitely many square roots if $\dim V > 1$.

Let R be a positive square root of T. We will prove that $Rv = \sqrt{\lambda}v$. This will imply that the behavior of R on the eigenvectors of T is uniquely determined. Because there is a basis of V consisting of eigenvectors of T (by the Spectral Theorem), this will imply that R is uniquely determined.

To prove that $Rv = \sqrt{\lambda}v$, note that the Spectral Theorem asserts that there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of R. Because R is a positive operator, all its eigenvalues are nonnegative. Thus there exist nonnegative numbers $\lambda_1, \dots, \lambda_n$ such that $Re_j = \sqrt{\lambda_j}e_j$ for $j = 1, \dots, n$.

Because $e_1, \dots, e_n$ is a basis of V, we can write

$$v = a_1e_1 + \cdots + a_ne_n$$

for some numbers $a_1, \dots, a_n \in \mathbf{F}$. Thus

$$Rv = a_1\sqrt{\lambda_1}e_1 + \cdots + a_n\sqrt{\lambda_n}e_n$$

and hence

$$R^2v = a_1\lambda_1e_1 + \cdots + a_n\lambda_ne_n.$$

But $R^2 = T$, and $Tv = \lambda v$. Thus the equation above implies

$$a_1\lambda e_1 + \cdots + a_n\lambda e_n = a_1\lambda_1e_1 + \cdots + a_n\lambda_ne_n.$$

The equation above implies that $a_j(\lambda - \lambda_j) = 0$ for $j = 1, \dots, n$. Hence

$$v = \sum_{\{j : \lambda_j = \lambda\}} a_je_j,$$

and thus

$$Rv = \sum_{\{j : \lambda_j = \lambda\}} a_j\sqrt{\lambda}e_j = \sqrt{\lambda}v,$$

as desired. ■

Isometries

Operators that preserve norms are sufficiently important to deserve a name:

7.37 Definition ***isometry***

- An operator $S \in \mathcal{L}(V)$ is called an ***isometry*** if
$$\|Sv\| = \|v\|$$
for all $v \in V$.
- In other words, an operator is an isometry if it preserves norms.

*The Greek word **isos** means equal; the Greek word **metron** means measure. Thus **isometry** literally means equal measure.*

For example, λI is an isometry whenever $\lambda \in \mathbf{F}$ satisfies $|\lambda| = 1$. We will see soon that if $\mathbf{F} = \mathbf{C}$, then the next example includes all isometries.

7.38 Example Suppose $\lambda_1, \ldots, \lambda_n$ are scalars with absolute value 1 and $S \in \mathcal{L}(V)$ satisfies $Se_j = \lambda_j e_j$ for some orthonormal basis $e_1, \ldots, e_n$ of V. Show that S is an isometry.

Solution Suppose $v \in V$. Then

7.39
$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

and

7.40
$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2,$$

where we have used 6.30. Applying S to both sides of 7.39 gives

$$\begin{aligned} Sv &= \langle v, e_1\rangle Se_1 + \cdots + \langle v, e_n\rangle Se_n \\ &= \lambda_1\langle v, e_1\rangle e_1 + \cdots + \lambda_n\langle v, e_n\rangle e_n. \end{aligned}$$

The last equation, along with the equation $|\lambda_j| = 1$, shows that

7.41
$$\|Sv\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2.$$

Comparing 7.40 and 7.41 shows that $\|v\| = \|Sv\|$. In other words, S is an isometry.

The next result provides several conditions that are equivalent to being an isometry. The equivalence of (a) and (b) shows that an operator is an isometry if and only if it preserves inner products. The equivalence of (a) and (c) [or (d)] shows that an operator is an isometry if and only if the list of columns of its matrix with respect to every [or some] basis is orthonormal. Exercise 10 implies that in the previous sentence we can replace "columns" with "rows".

An isometry on a real inner product space is often called an ***orthogonal*** *operator. An isometry on a complex inner product space is often called a* ***unitary*** *operator. We use the term isometry so that our results can apply to both real and complex inner product spaces.*

7.42 Characterization of isometries

Suppose $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \dots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \dots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $Se_1, \dots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

Proof First suppose (a) holds, so S is an isometry. Exercises 19 and 20 in Section 6.A show that inner products can be computed from norms. Because S preserves norms, this implies that S preserves inner products, and hence (b) holds. More precisely, if V is a real inner product space, then for every $u, v \in V$ we have

$$\begin{aligned}\langle Su, Sv \rangle &= (\|Su + Sv\|^2 - \|Su - Sv\|^2)/4 \\ &= (\|S(u+v)\|^2 - \|S(u-v)\|^2)/4 \\ &= (\|u+v\|^2 - \|u-v\|^2)/4 \\ &= \langle u, v \rangle,\end{aligned}$$

where the first equality comes from Exercise 19 in Section 6.A, the second equality comes from the linearity of S, the third equality holds because S is an isometry, and the last equality again comes from Exercise 19 in Section 6.A. If V is a complex inner product space, then use Exercise 20 in Section 6.A instead of Exercise 19 to obtain the same conclusion. In either case, we see that (b) holds.

Now suppose (b) holds, so S preserves inner products. Suppose that $e_1, \dots, e_n$ is an orthonormal list of vectors in V. Then we see that the list $Se_1, \dots, Se_n$ is orthonormal because $\langle Se_j, Se_k \rangle = \langle e_j, e_k \rangle$. Thus (c) holds.

Clearly (c) implies (d).

Now suppose (d) holds. Let $e_1, \dots, e_n$ be an orthonormal basis of V such that $Se_1, \dots, Se_n$ is orthonormal. Thus

$$\langle S^* S e_j, e_k \rangle = \langle e_j, e_k \rangle$$

for $j, k = 1, \dots, n$ [because the term on the left equals $\langle Se_j, Se_k \rangle$ and $(Se_1, \dots, Se_n)$ is orthonormal]. All vectors $u, v \in V$ can be written as linear combinations of $e_1, \dots, e_n$, and thus the equation above implies that $\langle S^* S u, v \rangle = \langle u, v \rangle$. Hence $S^* S = I$; in other words, (e) holds.

Now suppose (e) holds, so that $S^* S = I$. In general, an operator S need not commute with S^*. However, $S^* S = I$ if and only if $SS^* = I$; this is a special case of Exercise 10 in Section 3.D. Thus $SS^* = I$, showing that (f) holds.

Now suppose (f) holds, so $SS^* = I$. If $v \in V$, then

$$\|S^* v\|^2 = \langle S^* v, S^* v \rangle = \langle SS^* v, v \rangle = \langle v, v \rangle = \|v\|^2.$$

Thus S^* is an isometry, showing that (g) holds.

Now suppose (g) holds, so S^* is an isometry. We know that (a) $\Rightarrow$ (e) and (a) $\Rightarrow$ (f) because we have shown (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f). Using the implications (a) $\Rightarrow$ (e) and (a) $\Rightarrow$ (f) but with S replaced with S^* [and using the equation $(S^*)^* = S$], we conclude that $SS^* = I$ and $S^* S = I$. Thus S is invertible and $S^{-1} = S^*$; in other words, (h) holds.

Now suppose (h) holds, so S is invertible and $S^{-1} = S^*$. Thus $S^* S = I$. If $v \in V$, then

$$\|Sv\|^2 = \langle Sv, Sv \rangle = \langle S^* S v, v \rangle = \langle v, v \rangle = \|v\|^2.$$

Thus S is an isometry, showing that (a) holds.

We have shown (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (g) $\Rightarrow$ (h) $\Rightarrow$ (a), completing the proof. ■

The previous result shows that every isometry is normal [see (a), (e), and (f) of 7.42]. Thus characterizations of normal operators can be used to give descriptions of isometries. We do this in the next result in the complex case and in Chapter 9 in the real case (see 9.36).

7.43 Description of isometries when $\mathbf{F} = \mathbf{C}$

Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry.

(b) There is an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value 1.

Proof We have already shown (see Example 7.38) that (b) implies (a).

To prove the other direction, suppose (a) holds, so S is an isometry. By the Complex Spectral Theorem (7.24), there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of S. For $j \in \{1, \dots, n\}$, let λ_j be the eigenvalue corresponding to e_j. Then

$$|\lambda_j| = \|\lambda_j e_j\| = \|Se_j\| = \|e_j\| = 1.$$

Thus each eigenvalue of S has absolute value 1, completing the proof. ■

EXERCISES 7.C

1 Prove or give a counterexample: If $T \in \mathcal{L}(V)$ is self-adjoint and there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $\langle Te_j, e_j \rangle \geq 0$ for each j, then T is a positive operator.

2 Suppose T is a positive operator on V. Suppose $v, w \in V$ are such that

$$Tv = w \quad \text{and} \quad Tw = v.$$

Prove that $v = w$.

3 Suppose T is a positive operator on V and U is a subspace of V invariant under T. Prove that $T|_U \in \mathcal{L}(U)$ is a positive operator on U.

4 Suppose $T \in \mathcal{L}(V, W)$. Prove that T^*T is a positive operator on V and TT^* is a positive operator on W.

5 Prove that the sum of two positive operators on V is positive.

6 Suppose $T \in \mathcal{L}(V)$ is positive. Prove that T^k is positive for every positive integer k.

7 Suppose T is a positive operator on V. Prove that T is invertible if and only if

$$\langle Tv, v\rangle > 0$$

for every $v \in V$ with $v \neq 0$.

8 Suppose $T \in \mathcal{L}(V)$. For $u, v \in V$, define $\langle u, v\rangle_T$ by

$$\langle u, v\rangle_T = \langle Tu, v\rangle.$$

Prove that $\langle \cdot, \cdot\rangle_T$ is an inner product on V if and only if T is an invertible positive operator (with respect to the original inner product $\langle \cdot, \cdot\rangle$).

9 Prove or disprove: the identity operator on $\mathbf{F}^2$ has infinitely many self-adjoint square roots.

10 Suppose $S \in \mathcal{L}(V)$. Prove that the following are equivalent:

(a) S is an isometry;

(b) $\langle S^*u, S^*v\rangle = \langle u, v\rangle$ for all $u, v \in V$;

(c) $S^*e_1, \dots, S^*e_m$ is an orthonormal list for every orthonormal list of vectors $e_1, \dots, e_m$ in V;

(d) $S^*e_1, \dots, S^*e_n$ is an orthonormal basis for some orthonormal basis $e_1, \dots, e_n$ of V.

11 Suppose T_1, T_2 are normal operators on $\mathcal{L}(\mathbf{F}^3)$ and both operators have $2, 5, 7$ as eigenvalues. Prove that there exists an isometry $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T_1 = S^*T_2S$.

12 Give an example of two self-adjoint operators $T_1, T_2 \in \mathcal{L}(\mathbf{F}^4)$ such that the eigenvalues of both operators are $2, 5, 7$ but there does not exist an isometry $S \in \mathcal{L}(\mathbf{F}^4)$ such that $T_1 = S^*T_2S$. Be sure to explain why there is no isometry with the required property.

13 Prove or give a counterexample: if $S \in \mathcal{L}(V)$ and there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Se_j\| = 1$ for each e_j, then S is an isometry.

14 Let T be the second derivative operator in Exercise 21 in Section 7.A. Show that $-T$ is a positive operator.

7.D Polar Decomposition and Singular Value Decomposition

Polar Decomposition

Recall our analogy between $\mathbf{C}$ and $\mathcal{L}(V)$. Under this analogy, a complex number z corresponds to an operator T, and $\bar{z}$ corresponds to T^*. The real numbers $(z = \bar{z})$ correspond to the self-adjoint operators $(T = T^*)$, and the nonnegative numbers correspond to the (badly named) positive operators.

Another distinguished subset of $\mathbf{C}$ is the unit circle, which consists of the complex numbers z such that $|z| = 1$. The condition $|z| = 1$ is equivalent to the condition $\bar{z}z = 1$. Under our analogy, this would correspond to the condition $T^*T = I$, which is equivalent to T being an isometry (see 7.42). In other words, the unit circle in $\mathbf{C}$ corresponds to the isometries.

Continuing with our analogy, note that each complex number z except 0 can be written in the form

$$z = \left(\frac{z}{|z|}\right)|z| = \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z},$$

where the first factor, namely, $z/|z|$, is an element of the unit circle. Our analogy leads us to guess that each operator $T \in \mathcal{L}(V)$ can be written as an isometry times $\sqrt{T^*T}$. That guess is indeed correct, as we now prove after defining the obvious notation, which is justified by 7.36.

7.44 Notation $\sqrt{T}$

If T is a positive operator, then $\sqrt{T}$ denotes the unique positive square root of T.

Now we can state and prove the Polar Decomposition, which gives a beautiful description of an arbitrary operator on V. Note that T^*T is a positive operator for every $T \in \mathcal{L}(V)$, and thus $\sqrt{T^*T}$ is well defined.

7.45 Polar Decomposition

Suppose $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Proof If $v \in V$, then

$$\begin{aligned}\|Tv\|^2 = \langle Tv, Tv\rangle &= \langle T^*Tv, v\rangle \\ &= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \\ &= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \\ &= \|\sqrt{T^*T}v\|^2.\end{aligned}$$

Thus

7.46 $$\|Tv\| = \|\sqrt{T^*T}v\|$$

for all $v \in V$.

Define a linear map $S_1\colon \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$ by

7.47 $$S_1(\sqrt{T^*T}v) = Tv.$$

The idea of the proof is to extend S_1 to an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Now for the details.

First we must check that S_1 is well defined. To do this, suppose $v_1, v_2 \in V$ are such that $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$. For the definition given by 7.47 to make sense, we must show that $Tv_1 = Tv_2$. Note that

$$\begin{aligned}\|Tv_1 - Tv_2\| &= \|T(v_1 - v_2)\| \\ &= \|\sqrt{T^*T}(v_1 - v_2)\| \\ &= \|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\| \\ &= 0,\end{aligned}$$

where the second equality holds by 7.46. The equation above shows that $Tv_1 = Tv_2$, so S_1 is indeed well defined. You should verify that S_1 is a linear map.

We see from 7.47 that S_1 maps $\operatorname{range}\sqrt{T^*T}$ onto $\operatorname{range} T$. Clearly 7.46 and 7.47 imply that

$$\|S_1 u\| = \|u\|$$

for all $u \in \operatorname{range}\sqrt{T^*T}$.

The rest of the proof extends S_1 to an isometry S on all of V.

In particular, S_1 is injective. Thus from the Fundamental Theorem of Linear Maps (3.22), applied to S_1, we have

$$\dim \operatorname{range}\sqrt{T^*T} = \dim \operatorname{range} T.$$

This implies that $\dim(\operatorname{range}\sqrt{T^*T})^\perp = \dim(\operatorname{range} T)^\perp$ (see 6.50). Thus orthonormal bases $e_1, \ldots, e_m$ of $(\operatorname{range}\sqrt{T^*T})^\perp$ and $f_1, \ldots, f_m$ of $(\operatorname{range} T)^\perp$ can be chosen; the key point here is that these two orthonormal bases have the same length (denoted m). Now define a linear map $S_2\colon (\operatorname{range}\sqrt{T^*T})^\perp \to (\operatorname{range} T)^\perp$ by

$$S_2(a_1 e_1 + \cdots + a_m e_m) = a_1 f_1 + \cdots + a_m f_m.$$

For all $w \in (\operatorname{range}\sqrt{T^*T})^\perp$, we have $\|S_2 w\| = \|w\|$ (from 6.25).

Now let S be the operator on V that equals S_1 on $\operatorname{range}\sqrt{T^*T}$ and equals S_2 on $(\operatorname{range}\sqrt{T^*T})^\perp$. More precisely, recall that each $v \in V$ can be written uniquely in the form

7.48 $$v = u + w,$$

where $u \in \operatorname{range}\sqrt{T^*T}$ and $w \in (\operatorname{range}\sqrt{T^*T})^\perp$ (see 6.47). For $v \in V$ with decomposition as above, define Sv by

$$Sv = S_1 u + S_2 w.$$

For each $v \in V$ we have

$$S(\sqrt{T^*T}v) = S_1(\sqrt{T^*T}v) = Tv,$$

so $T = S\sqrt{T^*T}$, as desired. All that remains is to show that S is an isometry. However, this follows easily from two uses of the Pythagorean Theorem: if $v \in V$ has decomposition as in 7.48, then

$$\|Sv\|^2 = \|S_1 u + S_2 w\|^2 = \|S_1 u\|^2 + \|S_2 w\|^2 = \|u\|^2 + \|w\|^2 = \|v\|^2;$$

the second equality holds because $S_1 u \in \operatorname{range} T$ and $S_2 w \in (\operatorname{range} T)^\perp$. ■

The Polar Decomposition (7.45) states that each operator on V is the product of an isometry and a positive operator. Thus we can write each operator on V as the product of two operators, each of which comes from a class that we can completely describe and that we understand reasonably well. The isometries are described by 7.43 and 9.36; the positive operators are described by the Spectral Theorem (7.24 and 7.29).

Specifically, consider the case $\mathbf{F} = \mathbf{C}$, and suppose $T = S\sqrt{T^*T}$ is a Polar Decomposition of an operator $T \in \mathcal{L}(V)$, where S is an isometry. Then there is an orthonormal basis of V with respect to which S has a diagonal matrix, and there is an orthonormal basis of V with respect to which $\sqrt{T^*T}$ has a diagonal matrix. **Warning:** there may not exist an orthonormal basis that simultaneously puts the matrices of both S and $\sqrt{T^*T}$ into these nice diagonal forms. In other words, S may require one orthonormal basis and $\sqrt{T^*T}$ may require a different orthonormal basis.

Singular Value Decomposition

The eigenvalues of an operator tell us something about the behavior of the operator. Another collection of numbers, called the singular values, is also useful. Recall that eigenspaces and the notation E are defined in 5.36.

7.49 Definition ***singular values***

Suppose $T \in \mathcal{L}(V)$. The ***singular values*** of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim E(\lambda, \sqrt{T^*T})$ times.

The singular values of T are all nonnegative, because they are the eigenvalues of the positive operator $\sqrt{T^*T}$.

7.50 Example Define $T \in \mathcal{L}(\mathbf{F}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4).$$

Find the singular values of T.

Solution A calculation shows $T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4)$, as you should verify. Thus

$$\sqrt{T^*T}(z_1, z_2, z_3, z_4) = (3z_1, 2z_2, 0, 3z_4),$$

and we see that the eigenvalues of $\sqrt{T^*T}$ are $3, 2, 0$ and

$$\dim E(3, \sqrt{T^*T}) = 2, \ \dim E(2, \sqrt{T^*T}) = 1, \ \dim E(0, \sqrt{T^*T}) = 1.$$

Hence the singular values of T are $3, 3, 2, 0$.

Note that -3 and 0 are the only eigenvalues of T. Thus in this case, the collection of eigenvalues did not pick up the number 2 that appears in the definition (and hence the behavior) of T, but the collection of singular values does include 2.

Each $T \in \mathcal{L}(V)$ has $\dim V$ singular values, as can be seen by applying the Spectral Theorem and 5.41 [see especially part (e)] to the positive (hence self-adjoint) operator $\sqrt{T^*T}$. For example, the operator T defined in Example 7.50 on the four-dimensional vector space $\mathbf{F}^4$ has four singular values (they are $3, 3, 2, 0$), as we saw above.

The next result shows that every operator on V has a clean description in terms of its singular values and two orthonormal bases of V.

7.51 Singular Value Decomposition

Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \dots, s_n$. Then there exist orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ of V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$.

Proof By the Spectral Theorem applied to $\sqrt{T^*T}$, there is an orthonormal basis $e_1, \dots, e_n$ of V such that $\sqrt{T^*T}e_j = s_j e_j$ for $j = 1, \dots, n$.

We have

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

for every $v \in V$ (see 6.30). Apply $\sqrt{T^*T}$ to both sides of this equation, getting

$$\sqrt{T^*T}v = s_1\langle v, e_1\rangle e_1 + \cdots + s_n\langle v, e_n\rangle e_n$$

for every $v \in V$. By the Polar Decomposition (see 7.45), there is an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Apply S to both sides of the equation above, getting

$$Tv = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n$$

for every $v \in V$. For each j, let $f_j = Se_j$. Because S is an isometry, $f_1, \dots, f_n$ is an orthonormal basis of V (see 7.42). The equation above now becomes

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for every $v \in V$, completing the proof. ■

When we worked with linear maps from one vector space to a second vector space, we considered the matrix of a linear map with respect to a basis of the first vector space and a basis of the second vector space. When dealing with operators, which are linear maps from a vector space to itself, we almost always use only one basis, making it play both roles.

The Singular Value Decomposition allows us a rare opportunity to make good use of two different bases for the matrix of an operator. To do this, suppose $T \in \mathcal{L}(V)$. Let $s_1, \dots, s_n$ denote the singular values of T, and let $e_1, \dots, e_n$ and $f_1, \dots, f_n$ be orthonormal bases of V such that the Singular Value Decomposition 7.51 holds. Because $Te_j = s_j f_j$ for each j, we have

$$\mathcal{M}\big(T, (e_1, \dots, e_n), (f_1, \dots, f_n)\big) = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{pmatrix}.$$

In other words, every operator on V has a diagonal matrix with respect to some orthonormal bases of V, provided that we are permitted to use two different bases rather than a single basis as customary when working with operators.

Singular values and the Singular Value Decomposition have many applications (some are given in the exercises), including applications in computational linear algebra. To compute numeric approximations to the singular values of an operator T, first compute T^*T and then compute approximations to the eigenvalues of T^*T (good techniques exist for approximating eigenvalues of positive operators). The nonnegative square roots of these (approximate) eigenvalues of T^*T will be the (approximate) singular values of T. In other words, the singular values of T can be approximated without computing the square root of T^*T. The next result helps justify working with T^*T instead of $\sqrt{T^*T}$.

7.52 Singular values without taking square root of an operator

Suppose $T \in \mathcal{L}(V)$. Then the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, with each eigenvalue λ repeated $\dim E(\lambda, T^*T)$ times.

Proof The Spectral Theorem implies that there are an orthonormal basis $e_1, \dots, e_n$ and nonnegative numbers $\lambda_1, \dots, \lambda_n$ such that $T^*Te_j = \lambda_j e_j$ for $j = 1, \dots, n$. It is easy to see that $\sqrt{T^*T}e_j = \sqrt{\lambda_j}e_j$ for $j = 1, \dots, n$, which implies the desired result. ■

EXERCISES 7.D

1 Fix $u, x \in V$ with $u \neq 0$. Define $T \in \mathcal{L}(V)$ by

$$Tv = \langle v, u\rangle x$$

for every $v \in V$. Prove that

$$\sqrt{T^*T}v = \frac{\|x\|}{\|u\|}\langle v, u\rangle u$$

for every $v \in V$.

2 Give an example of $T \in \mathcal{L}(\mathbf{C}^2)$ such that 0 is the only eigenvalue of T and the singular values of T are $5, 0$.

3 Suppose $T \in \mathcal{L}(V)$. Prove that there exists an isometry $S \in \mathcal{L}(V)$ such that

$$T = \sqrt{TT^*}\, S.$$

4 Suppose $T \in \mathcal{L}(V)$ and s is a singular value of T. Prove that there exists a vector $v \in V$ such that $\|v\| = 1$ and $\|Tv\| = s$.

5 Suppose $T \in \mathcal{L}(\mathbf{C}^2)$ is defined by $T(x, y) = (-4y, x)$. Find the singular values of T.

6 Find the singular values of the differentiation operator $D \in \mathcal{P}(\mathbf{R}^2)$ defined by $Dp = p'$, where the inner product on $\mathcal{P}(\mathbf{R}^2)$ is as in Example 6.33.

7 Define $T \in \mathcal{L}(\mathbf{F}^3)$ by

$$T(z_1, z_2, z_3) = (z_3, 2z_1, 3z_2).$$

Find (explicitly) an isometry $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T = S\sqrt{T^*T}$.

8 Suppose $T \in \mathcal{L}(V)$, $S \in \mathcal{L}(V)$ is an isometry, and $R \in \mathcal{L}(V)$ is a positive operator such that $T = SR$. Prove that $R = \sqrt{T^*T}$.
[*The exercise above shows that if we write T as the product of an isometry and a positive operator (as in the Polar Decomposition 7.45), then the positive operator equals $\sqrt{T^*T}$.*]

9 Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if there exists a unique isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$.

10 Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Prove that the singular values of T equal the absolute values of the eigenvalues of T, repeated appropriately.

11 Suppose $T \in \mathcal{L}(V)$. Prove that T and T^* have the same singular values.

12 Prove or give a counterexample: if $T \in \mathcal{L}(V)$, then the singular values of T^2 equal the squares of the singular values of T.

13 Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if 0 is not a singular value of T.

14 Suppose $T \in \mathcal{L}(V)$. Prove that $\dim \operatorname{range} T$ equals the number of nonzero singular values of T.

15 Suppose $S \in \mathcal{L}(V)$. Prove that S is an isometry if and only if all the singular values of S equal 1.

16 Suppose $T_1, T_2 \in \mathcal{L}(V)$. Prove that T_1 and T_2 have the same singular values if and only if there exist isometries $S_1, S_2 \in \mathcal{L}(V)$ such that $T_1 = S_1 T_2 S_2$.

17 Suppose $T \in \mathcal{L}(V)$ has singular value decomposition given by

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n$$

for every $v \in V$, where $s_1, \ldots, s_n$ are the singular values of T and $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ are orthonormal bases of V.

(a) Prove that if $v \in V$, then

$$T^* v = s_1 \langle v, f_1 \rangle e_1 + \cdots + s_n \langle v, f_n \rangle e_n.$$

(b) Prove that if $v \in V$, then

$$T^* T v = {s_1}^2 \langle v, e_1 \rangle e_1 + \cdots + {s_n}^2 \langle v, e_n \rangle e_n.$$

(c) Prove that if $v \in V$, then

$$\sqrt{T^* T} v = s_1 \langle v, e_1 \rangle e_1 + \cdots + s_n \langle v, e_n \rangle e_n.$$

(d) Suppose T is invertible. Prove that if $v \in V$, then

$$T^{-1} v = \frac{\langle v, f_1 \rangle e_1}{s_1} + \cdots + \frac{\langle v, f_n \rangle e_n}{s_n}$$

for every $v \in V$.

18 Suppose $T \in \mathcal{L}(V)$. Let $\hat{s}$ denote the smallest singular value of T, and let s denote the largest singular value of T.

(a) Prove that $\hat{s}\|v\| \leq \|Tv\| \leq s\|v\|$ for every $v \in V$.

(b) Suppose λ is an eigenvalue of T. Prove that $\hat{s} \leq |\lambda| \leq s$.

19 Suppose $T \in \mathcal{L}(V)$. Show that T is uniformly continuous with respect to the metric d on V defined by $d(u, v) = \|u - v\|$.

20 Suppose $S, T \in \mathcal{L}(V)$. Let s denote the largest singular value of S, let t denote the largest singular value of T, and let r denote the largest singular value of $S + T$. Prove that $r \leq s + t$.

CHAPTER

8

Hypatia, the 5th century Egyptian mathematician and philosopher, as envisioned around 1900 by Alfred Seifert.

Operators on Complex Vector Spaces

In this chapter we delve deeper into the structure of operators, with most of the attention on complex vector spaces. An inner product does not help with this material, so we return to the general setting of a finite-dimensional vector space. To avoid some trivialities, we will assume that $V \neq \{0\}$. Thus our assumptions for this chapter are as follows:

8.1 Notation $\mathbf{F}$, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- generalized eigenvectors and generalized eigenspaces
- characteristic polynomial and the Cayley–Hamilton Theorem
- decomposition of an operator
- minimal polynomial
- Jordan Form

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_8

8.A Generalized Eigenvectors and Nilpotent Operators

Null Spaces of Powers of an Operator

We begin this chapter with a study of null spaces of powers of an operator.

8.2 Sequence of increasing null spaces

Suppose $T \in \mathcal{L}(V)$. Then

$$\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^k \subset \operatorname{null} T^{k+1} \subset \cdots.$$

Proof Suppose k is a nonnegative integer and $v \in \operatorname{null} T^k$. Then $T^k v = 0$, and hence $T^{k+1}v = T(T^k v) = T(0) = 0$. Thus $v \in \operatorname{null} T^{k+1}$. Hence $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$, as desired. ■

The next result says that if two consecutive terms in this sequence of subspaces are equal, then all later terms in the sequence are equal.

8.3 Equality in the sequence of null spaces

Suppose $T \in \mathcal{L}(V)$. Suppose m is a nonnegative integer such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Then

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \operatorname{null} T^{m+3} = \cdots.$$

Proof Let k be a positive integer. We want to prove that

$$\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}.$$

We already know from 8.2 that $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$.

To prove the inclusion in the other direction, suppose $v \in \operatorname{null} T^{m+k+1}$. Then

$$T^{m+1}(T^k v) = T^{m+k+1}v = 0.$$

Hence

$$T^k v \in \operatorname{null} T^{m+1} = \operatorname{null} T^m.$$

Thus $T^{m+k}v = T^m(T^k v) = 0$, which means that $v \in \operatorname{null} T^{m+k}$. This implies that $\operatorname{null} T^{m+k+1} \subset \operatorname{null} T^{m+k}$, completing the proof. ■

The proposition above raises the question of whether there exists a nonnegative integer m such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. The proposition below shows that this equality holds at least when m equals the dimension of the vector space on which T operates.

8.4 Null spaces stop growing

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \operatorname{null} T^{n+2} = \cdots.$$

Proof We need only prove that $\operatorname{null} T^n = \operatorname{null} T^{n+1}$ (by 8.3). Suppose this is not true. Then, by 8.2 and 8.3, we have

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^n \subsetneq \operatorname{null} T^{n+1},$$

where the symbol $\subsetneq$ means "contained in but not equal to". At each of the strict inclusions in the chain above, the dimension increases by at least 1. Thus $\dim \operatorname{null} T^{n+1} \geq n + 1$, a contradiction because a subspace of V cannot have a larger dimension than n. ■

Unfortunately, it is not true that $V = \operatorname{null} T \oplus \operatorname{range} T$ for each $T \in \mathcal{L}(V)$. However, the following result is a useful substitute.

8.5 V is the direct sum of $\operatorname{null} T^{\dim V}$ and $\operatorname{range} T^{\dim V}$

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n.$$

Proof First we show that

8.6 $$(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}.$$

Suppose $v \in (\operatorname{null} T^n) \cap (\operatorname{range} T^n)$. Then $T^n v = 0$, and there exists $u \in V$ such that $v = T^n u$. Applying T^n to both sides of the last equation shows that $T^n v = T^{2n} u$. Hence $T^{2n} u = 0$, which implies that $T^n u = 0$ (by 8.4). Thus $v = T^n u = 0$, completing the proof of 8.6.

Now 8.6 implies that $\operatorname{null} T^n + \operatorname{range} T^n$ is a direct sum (by 1.45). Also,

$$\dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) = \dim \operatorname{null} T^n + \dim \operatorname{range} T^n = \dim V,$$

where the first equality above comes from 3.78 and the second equality comes from the Fundamental Theorem of Linear Maps (3.22). The equation above implies that $\operatorname{null} T^n \oplus \operatorname{range} T^n = V$, as desired. ■

8.7 Example Suppose $T \in \mathcal{L}(\mathbf{F}^3)$ is defined by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

For this operator, $\operatorname{null} T + \operatorname{range} T$ is not a direct sum of subspaces, because $\operatorname{null} T = \{(z_1, 0, 0) : z_1 \in \mathbf{F}\}$ and $\operatorname{range} T = \{(z_1, 0, z_3) : z_1, z_3 \in \mathbf{F}\}$. Thus $\operatorname{null} T \cap \operatorname{range} T \neq \{0\}$ and hence $\operatorname{null} T + \operatorname{range} T$ is not a direct sum. Also note that $\operatorname{null} T + \operatorname{range} T \neq \mathbf{F}^3$.

However, we have $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$. Thus we see that $\operatorname{null} T^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{F}\}$ and $\operatorname{range} T^3 = \{(0, 0, z_3) : z_3 \in \mathbf{F}\}$. Hence $\mathbf{F}^3 = \operatorname{null} T^3 \oplus \operatorname{range} T^3$.

Generalized Eigenvectors

Unfortunately, some operators do not have enough eigenvectors to lead to a good description. Thus in this subsection we introduce the concept of generalized eigenvectors, which will play a major role in our description of the structure of an operator.

To understand why we need more than eigenvectors, let's examine the question of describing an operator by decomposing its domain into invariant subspaces. Fix $T \in \mathcal{L}(V)$. We seek to describe T by finding a "nice" direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a subspace of V invariant under T. The simplest possible nonzero invariant subspaces are 1-dimensional. A decomposition as above where each U_j is a 1-dimensional subspace of V invariant under T is possible if and only if V has a basis consisting of eigenvectors of T (see 5.41). This happens if and only if V has an eigenspace decomposition

8.8
$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (see 5.41).

The Spectral Theorem in the previous chapter shows that if V is an inner product space, then a decomposition of the form 8.8 holds for every normal operator if $\mathbf{F} = \mathbf{C}$ and for every self-adjoint operator if $\mathbf{F} = \mathbf{R}$ because operators of those types have enough eigenvectors to form a basis of V (see 7.24 and 7.29).

Sadly, a decomposition of the form 8.8 may not hold for more general operators, even on a complex vector space. An example was given by the operator in 5.43, which does not have enough eigenvectors for 8.8 to hold. Generalized eigenvectors and generalized eigenspaces, which we now introduce, will remedy this situation.

8.9 Definition ***generalized eigenvector***

Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a ***generalized eigenvector*** of T corresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0$$

for some positive integer j.

Although j is allowed to be an arbitrary integer in the equation

$$(T - \lambda I)^j v = 0$$

in the definition of a generalized eigenvector, we will soon prove that every generalized eigenvector satisfies this equation with $j = \dim V$.

Note that we do not define the concept of a generalized eigenvalue, because this would not lead to anything new. Reason: if $(T - \lambda I)^j$ is not injective for some positive integer j, then $T - \lambda I$ is not injective, and hence λ is an eigenvalue of T.

8.10 Definition ***generalized eigenspace,*** $G(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The ***generalized eigenspace*** of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Because every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the definition of generalized eigenvector), each eigenspace is contained in the corresponding generalized eigenspace. In other words, if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then

$$E(\lambda, T) \subset G(\lambda, T).$$

The next result implies that if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then $G(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V).

8.11 Description of generalized eigenspaces

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Then $G(\lambda, T) = \operatorname{null}(T - \lambda I)^{\dim V}$.

Proof Suppose $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. The definitions imply $v \in G(\lambda, T)$. Thus $G(\lambda, T) \supset \operatorname{null}(T - \lambda I)^{\dim V}$.

Conversely, suppose $v \in G(\lambda, T)$. Thus there is a positive integer j such that

$$v \in \operatorname{null}(T - \lambda I)^j.$$

From 8.2 and 8.4 (with $T - \lambda I$ replacing T), we get $v \in \operatorname{null}(T - \lambda I)^{\dim V}$. Thus $G(\lambda, T) \subset \operatorname{null}(T - \lambda I)^{\dim V}$, completing the proof. ■

8.12 **Example** Define $T \in \mathcal{L}(\mathbf{C}^3)$ by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

(a) Find all eigenvalues of T, the corresponding eigenspaces, and the corresponding generalized eigenspaces.

(b) Show that $\mathbf{C}^3$ is the direct sum of generalized eigenspaces corresponding to the distinct eigenvalues of T.

Solution

(a) A routine use of the definition of eigenvalue shows that the eigenvalues of T are 0 and 5. The corresponding eigenspaces are easily seen to be $E(0, T) = \{(z_1, 0, 0) : z_1 \in \mathbf{C}\}$ and $E(5, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}$.

Note that this operator T does not have enough eigenvectors to span its domain $\mathbf{C}^3$.

We have $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$ for all $z_1, z_2, z_3 \in \mathbf{C}$. Thus 8.11 implies that $G(0, T) = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbf{C}\}$.

We have $(T - 5I)^3(z_1, z_2, z_3) = (-125z_1 + 300z_2, -125z_2, 0)$. Thus 8.11 implies that $G(5, T) = \{(0, 0, z_3) : z_3 \in \mathbf{C}\}$.

(b) The results in part (a) show that $\mathbf{C}^3 = G(0, T) \oplus G(5, T)$.

One of our major goals in this chapter is to show that the result in part (b) of the example above holds in general for operators on finite-dimensional complex vector spaces; we will do this in 8.21.

We saw earlier (5.10) that eigenvectors corresponding to distinct eigenvalues are linearly independent. Now we prove a similar result for generalized eigenvectors.

8.13 Linearly independent generalized eigenvectors

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.

Proof Suppose $a_1, \dots, a_m$ are complex numbers such that

8.14 $$0 = a_1 v_1 + \cdots + a_m v_m.$$

Let k be the largest nonnegative integer such that $(T - \lambda_1 I)^k v_1 \neq 0$. Let

$$w = (T - \lambda_1 I)^k v_1.$$

Thus

$$(T - \lambda_1 I)w = (T - \lambda_1 I)^{k+1} w = 0,$$

and hence $Tw = \lambda_1 w$. Thus $(T - \lambda I)w = (\lambda_1 - \lambda)w$ for every $\lambda \in \mathbf{F}$ and hence

8.15 $$(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w$$

for every $\lambda \in \mathbf{F}$, where $n = \dim V$.

Apply the operator

$$(T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n$$

to both sides of 8.14, getting

$$\begin{aligned} 0 &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n v_1 \\ &= a_1 (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n w \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n w, \end{aligned}$$

where we have used 8.11 to get the first equation above and 8.15 to get the last equation above.

The equation above implies that $a_1 = 0$. In a similar fashion, $a_j = 0$ for each j, which implies that $v_1, \dots, v_m$ is linearly independent. ■

Nilpotent Operators

8.16 Definition ***nilpotent***

An operator is called ***nilpotent*** if some power of it equals 0.

8.17 Example ***nilpotent operators***

(a) The operator $N \in \mathcal{L}(\mathbf{F}^4)$ defined by

$$N(z_1, z_2, z_3, z_4) = (z_3, z_4, 0, 0)$$

is nilpotent because $N^2 = 0$.

(b) The operator of differentiation on $\mathcal{P}_m(\mathbf{R})$ is nilpotent because the $(m+1)^{\text{st}}$ derivative of every polynomial of degree at most m equals 0. Note that on this space of dimension $m+1$, we need to raise the nilpotent operator to the power $m+1$ to get the 0 operator.

*The Latin word **nil** means nothing or zero; the Latin word **potent** means power. Thus **nilpotent** literally means zero power.*

The next result shows that we never need to use a power higher than the dimension of the space.

8.18 Nilpotent operator raised to dimension of domain is 0

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $N^{\dim V} = 0$.

Proof Because N is nilpotent, $G(0, N) = V$. Thus 8.11 implies that $\operatorname{null} N^{\dim V} = V$, as desired. ∎

Given an operator T on V, we want to find a basis of V such that the matrix of T with respect to this basis is as simple as possible, meaning that the matrix contains many 0's.

If V is a complex vector space, a proof of the next result follows easily from Exercise 7, 5.27, and 5.32. But the proof given here uses simpler ideas than needed to prove 5.27, and it works for both real and complex vector spaces.

The next result shows that if N is nilpotent, then we can choose a basis of V such that the matrix of N with respect to this basis has more than half of its entries equal to 0. Later in this chapter we will do even better.

8.19 Matrix of a nilpotent operator

Suppose N is a nilpotent operator on V. Then there is a basis of V with respect to which the matrix of N has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix};$$

here all entries on and below the diagonal are 0's.

Proof First choose a basis of null N. Then extend this to a basis of null N^2. Then extend to a basis of null N^3. Continue in this fashion, eventually getting a basis of V (because 8.18 states that null $N^{\dim V} = V$).

Now let's think about the matrix of N with respect to this basis. The first column, and perhaps additional columns at the beginning, consists of all 0's, because the corresponding basis vectors are in null N. The next set of columns comes from basis vectors in null N^2. Applying N to any such vector, we get a vector in null N; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus all nonzero entries in these columns lie above the diagonal. The next set of columns comes from basis vectors in null N^3. Applying N to any such vector, we get a vector in null N^2; in other words, we get a vector that is a linear combination of the previous basis vectors. Thus once again, all nonzero entries in these columns lie above the diagonal. Continue in this fashion to complete the proof. ■

EXERCISES 8.A

1 Define $T \in \mathcal{L}(\mathbf{C}^2)$ by

$$T(w, z) = (z, 0).$$

Find all generalized eigenvectors of T.

2 Define $T \in \mathcal{L}(\mathbf{C}^2)$ by

$$T(w, z) = (-z, w).$$

Find the generalized eigenspaces corresponding to the distinct eigenvalues of T.

3 Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that $G(\lambda, T) = G(\frac{1}{\lambda}, T^{-1})$ for every $\lambda \in \mathbf{F}$ with $\lambda \neq 0$.

4 Suppose $T \in \mathcal{L}(V)$ and $\alpha, \beta \in \mathbf{F}$ with $\alpha \neq \beta$. Prove that

$$G(\alpha, T) \cap G(\beta, T) = \{0\}.$$

5 Suppose $T \in \mathcal{L}(V)$, m is a positive integer, and $v \in V$ is such that $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that

$$v, Tv, T^2 v, \ldots, T^{m-1}v$$

is linearly independent.

6 Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by $T(z_1, z_2, z_3) = (z_2, z_3, 0)$. Prove that T has no square root. More precisely, prove that there does not exist $S \in \mathcal{L}(\mathbf{C}^3)$ such that $S^2 = T$.

7 Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that 0 is the only eigenvalue of N.

8 Prove or give a counterexample: The set of nilpotent operators on V is a subspace of $\mathcal{L}(V)$.

9 Suppose $S, T \in \mathcal{L}(V)$ and ST is nilpotent. Prove that TS is nilpotent.

10 Suppose that $T \in \mathcal{L}(V)$ is not nilpotent. Let $n = \dim V$. Show that $V = \operatorname{null} T^{n-1} \oplus \operatorname{range} T^{n-1}$.

11 Prove or give a counterexample: If V is a complex vector space and $\dim V = n$ and $T \in \mathcal{L}(V)$, then T^n is diagonalizable.

12 Suppose $N \in \mathcal{L}(V)$ and there exists a basis of V with respect to which N has an upper-triangular matrix with only 0's on the diagonal. Prove that N is nilpotent.

13 Suppose V is an inner product space and $N \in \mathcal{L}(V)$ is normal and nilpotent. Prove that $N = 0$.

14 Suppose V is an inner product space and $N \in \mathcal{L}(V)$ is nilpotent. Prove that there exists an orthonormal basis of V with respect to which N has an upper-triangular matrix.
[*If* $\mathbf{F} = \mathbf{C}$, *then the result above follows from Schur's Theorem* (6.38) *without the hypothesis that N is nilpotent. Thus the exercise above needs to be proved only when* $\mathbf{F} = \mathbf{R}$.]

15 Suppose $N \in \mathcal{L}(V)$ is such that $\operatorname{null} N^{\dim V - 1} \neq \operatorname{null} N^{\dim V}$. Prove that N is nilpotent and that

$$\dim \operatorname{null} N^j = j$$

for every integer j with $0 \leq j \leq \dim V$.

16 Suppose $T \in \mathcal{L}(V)$. Show that

$$V = \operatorname{range} T^0 \supset \operatorname{range} T^1 \supset \cdots \supset \operatorname{range} T^k \supset \operatorname{range} T^{k+1} \supset \cdots.$$

17 Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer such that

$$\operatorname{range} T^m = \operatorname{range} T^{m+1}.$$

Prove that $\operatorname{range} T^k = \operatorname{range} T^m$ for all $k > m$.

18 Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Prove that

$$\operatorname{range} T^n = \operatorname{range} T^{n+1} = \operatorname{range} T^{n+2} = \cdots.$$

19 Suppose $T \in \mathcal{L}(V)$ and m is a nonnegative integer. Prove that

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} \quad \text{if and only if} \quad \operatorname{range} T^m = \operatorname{range} T^{m+1}.$$

20 Suppose $T \in \mathcal{L}(\mathbf{C}^5)$ is such that $\operatorname{range} T^4 \neq \operatorname{range} T^5$. Prove that T is nilpotent.

21 Find a vector space W and $T \in \mathcal{L}(W)$ such that $\operatorname{null} T^k \subsetneq \operatorname{null} T^{k+1}$ and $\operatorname{range} T^k \supsetneq \operatorname{range} T^{k+1}$ for every positive integer k.

8.B *Decomposition of an Operator*

Description of Operators on Complex Vector Spaces

We saw earlier that the domain of an operator might not decompose into eigenspaces, even on a finite-dimensional complex vector space. In this section we will see that every operator on a finite-dimensional complex vector space has enough generalized eigenvectors to provide a decomposition.

We observed earlier that if $T \in \mathcal{L}(V)$, then null T and range T are invariant under T [see 5.3, parts (c) and (d)]. Now we show that the null space and the range of each polynomial of T is also invariant under T.

8.20 The null space and range of $p(T)$ are invariant under T

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$. Then null $p(T)$ and range $p(T)$ are invariant under T.

Proof Suppose $v \in \operatorname{null} p(T)$. Then $p(T)v = 0$. Thus

$$\big((p(T)\big)(Tv) = T\big(p(T)v\big) = T(0) = 0.$$

Hence $Tv \in \operatorname{null} p(T)$. Thus null $p(T)$ is invariant under T, as desired.

Suppose $v \in \operatorname{range} p(T)$. Then there exists $u \in V$ such that $v = p(T)u$. Thus

$$Tv = T\big(p(T)u\big) = p(T)(Tu).$$

Hence $Tv \in \operatorname{range} p(T)$. Thus range $p(T)$ is invariant under T, as desired. ■

The following major result shows that every operator on a complex vector space can be thought of as composed of pieces, each of which is a nilpotent operator plus a scalar multiple of the identity. Actually we have already done the hard work in our discussion of the generalized eigenspaces $G(\lambda, T)$, so at this point the proof is easy.

8.21 Description of operators on complex vector spaces

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(b) each $G(\lambda_j, T)$ is invariant under T;

(c) each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

Proof Let $n = \dim V$. Recall that $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^n$ for each j (by 8.11). From 8.20 [with $p(z) = (z - \lambda_j)^n$], we get (b). Obviously (c) follows from the definitions.

We will prove (a) by induction on n. To get started, note that the desired result holds if $n = 1$. Thus we can assume that $n > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Because V is a complex vector space, T has an eigenvalue (see 5.21); thus $m \geq 1$. Applying 8.5 to $T - \lambda_1 I$ shows that

8.22 $$V = G(\lambda_1, T) \oplus U,$$

where $U = \operatorname{range}(T - \lambda_1 I)^n$. Using 8.20 [with $p(z) = (z - \lambda_1)^n$], we see that U is invariant under T. Because $G(\lambda_1, T) \neq \{0\}$, we have $\dim U < n$. Thus we can apply our induction hypothesis to $T|_U$.

None of the generalized eigenvectors of $T|_U$ correspond to the eigenvalue λ_1, because all generalized eigenvectors of T corresponding to λ_1 are in $G(\lambda_1, T)$. Thus each eigenvalue of $T|_U$ is in $\{\lambda_2, \ldots, \lambda_m\}$.

By our induction hypothesis, $U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U)$. Combining this information with 8.22 will complete the proof if we can show that $G(\lambda_k, T|_U) = G(\lambda_k, T)$ for $k = 2, \ldots, m$.

Thus fix $k \in \{2, \ldots, m\}$. The inclusion $G(\lambda_k, T|_U) \subset G(\lambda_k, T)$ is clear.

To prove the inclusion in the other direction, suppose $v \in G(\lambda_k, T)$. By 8.22, we can write $v = v_1 + u$, where $v_1 \in G(\lambda_1, T)$ and $u \in U$. Our induction hypothesis implies that

$$u = v_2 + \cdots + v_m,$$

where each v_j is in $G(\lambda_j, T|_U)$, which is a subset of $G(\lambda_j, T)$. Thus

$$v = v_1 + v_2 + \cdots + v_m,$$

Because generalized eigenvectors corresponding to distinct eigenvalues are linearly independent (see 8.13), the equation above implies that each v_j equals 0 except possibly when $j = k$. In particular, $v_1 = 0$ and thus $v = u \in U$. Because $v \in U$, we can conclude that $v \in G(\lambda_k, T|_U)$, completing the proof. ■

As we know, an operator on a complex vector space may not have enough eigenvectors to form a basis of the domain. The next result shows that on a complex vector space there are enough generalized eigenvectors to do this.

8.23 A basis of generalized eigenvectors

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.

Proof Choose a basis of each $G(\lambda_j, T)$ in 8.21. Put all these bases together to form a basis of V consisting of generalized eigenvectors of T. ■

Multiplicity of an Eigenvalue

If V is a complex vector space and $T \in \mathcal{L}(V)$, then the decomposition of V provided by 8.21 can be a powerful tool. The dimensions of the subspaces involved in this decomposition are sufficiently important to get a name.

8.24 Definition ***multiplicity***

- Suppose $T \in \mathcal{L}(V)$. The ***multiplicity*** of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$.
- In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$.

The second bullet point above is justified by 8.11.

8.25 Example Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

The eigenvalues of T are 6 and 7, as follows from 5.32. You can verify that the generalized eigenspaces of T are as follows:

$$G(6, T) = \operatorname{span}\big((1, 0, 0), (0, 1, 0)\big) \quad \text{and} \quad G(7, T) = \operatorname{span}\big((10, 2, 1)\big).$$

Thus the eigenvalue 6 has multiplicity 2 and the eigenvalue 7 has multiplicity 1.

The direct sum $\mathbf{C}^3 = G(6, T) \oplus G(7, T)$ is the decomposition promised by 8.21. A basis of $\mathbf{C}^3$ consisting of generalized eigenvectors of T, as promised by 8.23, is

$$(1, 0, 0), (0, 1, 0), (10, 2, 1).$$

In Example 8.25, the sum of the multiplicities of the eigenvalues of T equals 3, which is the dimension of the domain of T. The next result shows that this always happens on a complex vector space.

8.26 Sum of the multiplicities equals $\dim V$

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then the sum of the multiplicities of all the eigenvalues of T equals $\dim V$.

Proof The desired result follows from 8.21 and the obvious formula for the dimension of a direct sum (see 3.78 or Exercise 16 in Section 2.C). ■

The terms ***algebraic multiplicity*** and ***geometric multiplicity*** are used in some books. In case you encounter this terminology, be aware that the algebraic multiplicity is the same as the multiplicity defined here and the geometric multiplicity is the dimension of the corresponding eigenspace. In other words, if $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T, then

$$\begin{aligned}\text{algebraic multiplicity of } \lambda &= \dim \operatorname{null}(T - \lambda I)^{\dim V} = \dim G(\lambda, T),\\ \text{geometric multiplicity of } \lambda &= \dim \operatorname{null}(T - \lambda I) = \dim E(\lambda, T).\end{aligned}$$

Note that as defined above, the algebraic multiplicity also has a geometric meaning as the dimension of a certain null space. The definition of multiplicity given here is cleaner than the traditional definition that involves determinants; 10.25 implies that these definitions are equivalent.

Block Diagonal Matrices

To interpret our results in matrix form, we make the following definition, generalizing the notion of a diagonal matrix.

Often we can understand a matrix better by thinking of it as composed of smaller matrices.

If each matrix A_j in the definition below is a 1-by-1 matrix, then we actually have a diagonal matrix.

8.27 Definition ***block diagonal matrix***

A ***block diagonal matrix*** is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \dots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

8.28 Example The 5-by-5 matrix

$$A = \begin{pmatrix} \begin{pmatrix} 4 \end{pmatrix} & 0 \quad 0 & 0 \quad 0 \\ \begin{matrix} 0 \\ 0 \end{matrix} & \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} \\ \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix} \end{pmatrix}$$

is a block diagonal matrix with

$$A = \begin{pmatrix} A_1 & & 0 \\ & A_2 & \\ 0 & & A_3 \end{pmatrix},$$

where

$$A_1 = \begin{pmatrix} 4 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix}.$$

Here the inner matrices in the 5-by-5 matrix above are blocked off to show how we can think of it as a block diagonal matrix.

Note that in the next result we get many more zeros in the matrix of T than are needed to make it upper triangular.

8.29 Block diagonal matrix with upper-triangular blocks

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \dots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_j is a d_j-by-d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

Proof Each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent [see 8.21(c)]. For each j, choose a basis of $G(\lambda_j, T)$, which is a vector space with dimension d_j, such that the matrix of $(T - \lambda_j I)|_{G(\lambda_j, T)}$ with respect to this basis is as in 8.19. Thus the matrix of $T|_{G(\lambda_j, T)}$, which equals $(T - \lambda_j I)|_{G(\lambda_j, T)} + \lambda_j I|_{G(\lambda_j, T)}$, with respect to this basis will look like the desired form shown above for A_j.

Putting the bases of the $G(\lambda_j, T)$'s together gives a basis of V [by 8.21(a)]. The matrix of T with respect to this basis has the desired form. ■

The 5-by-5 matrix in 8.28 is of the form promised by 8.29, with each of the blocks itself an upper-triangular matrix that is constant along the diagonal of the block. If T is an operator on a 5-dimensional vector space whose matrix is as in 8.28, then the eigenvalues of T are 4, 2, 1 (as follows from 5.32), with multiplicities 1, 2, 2.

8.30 **Example** Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix},$$

which is an upper-triangular matrix but is not of the form promised by 8.29.

As we saw in Example 8.25, the eigenvalues of T are 6 and 7 and the corresponding generalized eigenspaces are

$$G(6, T) = \operatorname{span}\big((1, 0, 0), (0, 1, 0)\big) \quad \text{and} \quad G(7, T) = \operatorname{span}\big((10, 2, 1)\big).$$

We also saw that a basis of $\mathbf{C}^3$ consisting of generalized eigenvectors of T is

$$(1, 0, 0), (0, 1, 0), (10, 2, 1).$$

The matrix of T with respect to this basis is

$$\left(\begin{array}{cc} \begin{pmatrix} 6 & 3 \\ 0 & 6 \end{pmatrix} & \begin{array}{c} 0 \\ 0 \end{array} \\ \begin{array}{cc} 0 & 0 \end{array} & \begin{pmatrix} 7 \end{pmatrix} \end{array}\right),$$

which is a matrix of the block diagonal form promised by 8.29.

When we discuss the Jordan Form in Section 8.D, we will see that we can find a basis with respect to which an operator T has a matrix with even more 0's than promised by 8.29. However, 8.29 and its equivalent companion 8.21 are already quite powerful. For example, in the next subsection we will use 8.21 to show that every invertible operator on a complex vector space has a square root.

Square Roots

Recall that a square root of an operator $T \in \mathcal{L}(V)$ is an operator $R \in \mathcal{L}(V)$ such that $R^2 = T$ (see 7.33). Every complex number has a square root, but not every operator on a complex vector space has a square root. For example, the operator on $\mathbf{C}^3$ in Exercise 6 in Section 8.A has no square root. The noninvertibility of that operator is no accident, as we will soon see. We begin by showing that the identity plus any nilpotent operator has a square root.

8.31 Identity plus nilpotent has a square root

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.

Proof Consider the Taylor series for the function $\sqrt{1+x}$:

8.32 $$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots.$$

Because $a_1 = 1/2$, the formula above shows that $1 + x/2$ is a good estimate for $\sqrt{1+x}$ when x is small.

We will not find an explicit formula for the coefficients or worry about whether the infinite sum converges because we will use this equation only as motivation.

Because N is nilpotent, $N^m = 0$ for some positive integer m. In 8.32, suppose we replace x with N and 1 with I. Then the infinite sum on the right side becomes a finite sum (because $N^j = 0$ for all $j \geq m$). In other words, we guess that there is a square root of $I + N$ of the form

$$I + a_1 N + a_2 N^2 + \cdots + a_{m-1} N^{m-1}.$$

Having made this guess, we can try to choose $a_1, a_2, \ldots, a_{m-1}$ such that the operator above has its square equal to $I + N$. Now

$$\begin{aligned}(I + a_1 N &+ a_2 N^2 + a_3 N^3 + \cdots + a_{m-1} N^{m-1})^2 \\ &= I + 2a_1 N + (2a_2 + a_1{}^2) N^2 + (2a_3 + 2a_1 a_2) N^3 + \cdots \\ &\quad + (2a_{m-1} + \text{terms involving } a_1, \ldots, a_{m-2}) N^{m-1}.\end{aligned}$$

We want the right side of the equation above to equal $I + N$. Hence choose a_1 such that $2a_1 = 1$ (thus $a_1 = 1/2$). Next, choose a_2 such that $2a_2 + a_1{}^2 = 0$ (thus $a_2 = -1/8$). Then choose a_3 such that the coefficient of N^3 on the right side of the equation above equals 0 (thus $a_3 = 1/16$). Continue in this fashion for $j = 4, \ldots, m-1$, at each step solving for a_j so that the coefficient of N^j on the right side of the equation above equals 0. Actually we do not care about the explicit formula for the a_j's. We need only know that some choice of the a_j's gives a square root of $I + N$. ■

The previous lemma is valid on real and complex vector spaces. However, the next result holds only on complex vector spaces. For example, the operator of multiplication by -1 on the 1-dimensional real vector space $\mathbf{R}$ has no square root.

8.33 Over $\mathbf{C}$, invertible operators have square roots

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Then T has a square root.

Proof Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. For each j, there exists a nilpotent operator $N_j \in \mathcal{L}\big(G(\lambda_j, T)\big)$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$ [see 8.21(c)]. Because T is invertible, none of the λ_j's equals 0, so we can write

$$T|_{G(\lambda_j,T)} = \lambda_j\Big(I + \frac{N_j}{\lambda_j}\Big)$$

for each j. Clearly N_j/λ_j is nilpotent, and so $I + N_j/\lambda_j$ has a square root (by 8.31). Multiplying a square root of the complex number λ_j by a square root of $I + N_j/\lambda_j$, we obtain a square root R_j of $T|_{G(\lambda_j,T)}$.

A typical vector $v \in V$ can be written uniquely in the form

$$v = u_1 + \cdots + u_m,$$

where each u_j is in $G(\lambda_j, T)$ (see 8.21). Using this decomposition, define an operator $R \in \mathcal{L}(V)$ by

$$Rv = R_1 u_1 + \cdots + R_m u_m.$$

You should verify that this operator R is a square root of T, completing the proof. ■

By imitating the techniques in this section, you should be able to prove that if V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible, then T has a k^{th} root for every positive integer k.

EXERCISES 8.B

1 Suppose V is a complex vector space, $N \in \mathcal{L}(V)$, and 0 is the only eigenvalue of N. Prove that N is nilpotent.

2 Give an example of an operator T on a finite-dimensional real vector space such that 0 is the only eigenvalue of T but T is not nilpotent.

3 Suppose $T \in \mathcal{L}(V)$. Suppose $S \in \mathcal{L}(V)$ is invertible. Prove that T and $S^{-1}TS$ have the same eigenvalues with the same multiplicities.

4 Suppose V is an n-dimensional complex vector space and T is an operator on V such that $\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}$. Prove that T has at most two distinct eigenvalues.

5 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that V has a basis consisting of eigenvectors of T if and only if every generalized eigenvector of T is an eigenvector of T.
[*For* $\mathbf{F} = \mathbf{C}$*, the exercise above adds an equivalence to the list in 5.41.*]

6 Define $N \in \mathcal{L}(\mathbf{F}^5)$ by

$$N(x_1, x_2, x_3, x_4, x_5) = (2x_2, 3x_3, -x_4, 4x_5, 0).$$

Find a square root of $I + N$.

7 Suppose V is a complex vector space. Prove that every invertible operator on V has a cube root.

8 Suppose $T \in \mathcal{L}(V)$ and 3 and 8 are eigenvalues of T. Let $n = \dim V$. Prove that $V = (\operatorname{null} T^{n-2}) \oplus (\operatorname{range} T^{n-2})$.

9 Suppose A and B are block diagonal matrices of the form

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix},$$

where A_j has the same size as B_j for $j = 1, \ldots, m$. Show that AB is a block diagonal matrix of the form

$$AB = \begin{pmatrix} A_1B_1 & & 0 \\ & \ddots & \\ 0 & & A_mB_m \end{pmatrix}.$$

10 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that there exist $D, N \in \mathcal{L}(V)$ such that $T = D + N$, the operator D is diagonalizable, N is nilpotent, and $DN = ND$.

11 Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Prove that for every basis of V with respect to which T has an upper-triangular matrix, the number of times that λ appears on the diagonal of the matrix of T equals the multiplicity of λ as an eigenvalue of T.

8.C Characteristic and Minimal Polynomials

The Cayley–Hamilton Theorem

The next definition associates a polynomial with each operator on V if $\mathbf{F} = \mathbf{C}$. For $\mathbf{F} = \mathbf{R}$, the corresponding definition will be given in the next chapter.

8.34 Definition ***characteristic polynomial***

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \dots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the ***characteristic polynomial*** of T.

8.35 Example Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is defined as in Example 8.25. Because the eigenvalues of T are 6, with multiplicity 2, and 7, with multiplicity 1, we see that the characteristic polynomial of T is $(z - 6)^2(z - 7)$.

8.36 Degree and zeros of characteristic polynomial

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then

(a) the characteristic polynomial of T has degree $\dim V$;

(b) the zeros of the characteristic polynomial of T are the eigenvalues of T.

Proof Clearly part (a) follows from 8.26 and part (b) follows from the definition of the characteristic polynomial. ■

Most texts define the characteristic polynomial using determinants (the two definitions are equivalent by 10.25). The approach taken here, which is considerably simpler, leads to the following easy proof of the Cayley–Hamilton Theorem. In the next chapter, we will see that this result also holds on real vector spaces (see 9.24).

8.37 Cayley–Hamilton Theorem

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

English mathematician Arthur Cayley (1821–1895) published three math papers before completing his undergraduate degree in 1842. Irish mathematician William Rowan Hamilton (1805–1865) was made a professor in 1827 when he was 22 years old and still an undergraduate!

Proof Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of the operator T, and let $d_1, \dots, d_m$ be the dimensions of the corresponding generalized eigenspaces $G(\lambda_1, T), \dots, G(\lambda_m, T)$. For each $j \in \{1, \dots, m\}$, we know that $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus we have $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0$ (by 8.18).

Every vector in V is a sum of vectors in $G(\lambda_1, T), \dots, G(\lambda_m, T)$ (by 8.21). Thus to prove that $q(T) = 0$, we need only show that $q(T)|_{G(\lambda_j, T)} = 0$ for each j.

Thus fix $j \in \{1, \dots, m\}$. We have

$$q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m}.$$

The operators on the right side of the equation above all commute, so we can move the factor $(T - \lambda_j I)^{d_j}$ to be the last term in the expression on the right. Because $(T - \lambda_j I)^{d_j}|_{G(\lambda_j, T)} = 0$, we conclude that $q(T)|_{G(\lambda_j, T)} = 0$, as desired. ■

The Minimal Polynomial

In this subsection we introduce another important polynomial associated with each operator. We begin with the following definition.

8.38 Definition ***monic polynomial***

A ***monic polynomial*** is a polynomial whose highest-degree coefficient equals 1.

8.39 Example The polynomial $2 + 9z^2 + z^7$ is a monic polynomial of degree 7.

8.40 Minimal polynomial

Suppose $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$.

Proof Let $n = \dim V$. Then the list

$$I, T, T^2, \dots, T^{n^2}$$

is not linearly independent in $\mathcal{L}(V)$, because the vector space $\mathcal{L}(V)$ has dimension n^2 (see 3.61) and we have a list of length $n^2 + 1$. Let m be the smallest positive integer such that the list

8.41 $$I, T, T^2, \dots, T^m$$

is linearly dependent. The Linear Dependence Lemma (2.21) implies that one of the operators in the list above is a linear combination of the previous ones. Because m was chosen to be the smallest positive integer such that the list above is linearly dependent, we conclude that T^m is a linear combination of $I, T, T^2, \dots, T^{m-1}$. Thus there exist scalars $a_0, a_1, a_2, \dots, a_{m-1} \in \mathbf{F}$ such that

8.42 $$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0.$$

Define a monic polynomial $p \in \mathcal{P}(\mathbf{F})$ by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m.$$

Then 8.42 implies that $p(T) = 0$.

To prove the uniqueness part of the result, note that the choice of m implies that no monic polynomial $q \in \mathcal{P}(\mathbf{F})$ with degree smaller than m can satisfy $q(T) = 0$. Suppose $q \in \mathcal{P}(\mathbf{F})$ is a monic polynomial with degree m and $q(T) = 0$. Then $(p - q)(T) = 0$ and $\deg(p - q) < m$. The choice of m now implies that $q = p$, completing the proof. ∎

The last result justifies the following definition.

8.43 Definition ***minimal polynomial***

Suppose $T \in \mathcal{L}(V)$. Then the ***minimal polynomial*** of T is the unique monic polynomial p of smallest degree such that $p(T) = 0$.

The proof of the last result shows that the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$. The Cayley–Hamilton Theorem (8.37) tells us that if V is a complex vector space, then the minimal polynomial of each operator on V has degree at most $\dim V$. This remarkable improvement also holds on real vector spaces, as we will see in the next chapter.

Suppose you are given the matrix (with respect to some basis) of an operator $T \in \mathcal{L}(V)$. You could program a computer to find the minimal polynomial of T as follows: Consider the system of linear equations

8.44 $$a_0\mathcal{M}(I) + a_1\mathcal{M}(T) + \cdots + a_{m-1}\mathcal{M}(T)^{m-1} = -\mathcal{M}(T)^m$$

Think of this as a system of $(\dim V)^2$ *linear equations in* m *variables* $a_0, a_1, \ldots, a_{m-1}$.

for successive values of $m = 1, 2, \ldots$ until this system of equations has a solution $a_0, a_1, a_2, \ldots, a_{m-1}$. The scalars $a_0, a_1, a_2, \ldots, a_{m-1}, 1$ will then be the coefficients of the minimal polynomial of T. All this can be computed using a familiar and fast (for a computer) process such as Gaussian elimination.

8.45 **Example** Let T be the operator on $\mathbf{C}^5$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Find the minimal polynomial of T.

Solution Because of the large number of 0's in this matrix, Gaussian elimination is not needed here. Simply compute powers of $\mathcal{M}(T)$, and then you will notice that there is clearly no solution to 8.44 until $m = 5$. Do the computations and you will see that the minimal polynomial of T equals $z^5 - 6z + 3$.

The next result completely characterizes the polynomials that when applied to an operator give the 0 operator.

8.46 $q(T) = 0$ implies q is a multiple of the minimal polynomial

Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.

Proof Let p denote the minimal polynomial of T.

First we prove the easy direction. Suppose q is a polynomial multiple of p. Thus there exists a polynomial $s \in \mathcal{P}(\mathbf{F})$ such that $q = ps$. We have

$$q(T) = p(T)s(T) = 0\, s(T) = 0,$$

as desired.

To prove the other direction, now suppose $q(T) = 0$. By the Division Algorithm for Polynomials (4.8), there exist polynomials $s, r \in \mathcal{P}(\mathbf{F})$ such that

8.47
$$q = ps + r$$

and $\deg r < \deg p$. We have

$$0 = q(T) = p(T)s(T) + r(T) = r(T).$$

The equation above implies that $r = 0$ (otherwise, dividing r by its highest-degree coefficient would produce a monic polynomial that when applied to T gives 0; this polynomial would have a smaller degree than the minimal polynomial, which would be a contradiction). Thus 8.47 becomes the equation $q = ps$. Hence q is a polynomial multiple of p, as desired. ■

The next result is stated only for complex vector spaces, because we have not yet defined the characteristic polynomial when $\mathbf{F} = \mathbf{R}$. However, the result also holds for real vector spaces, as we will see in the next chapter.

8.48 Characteristic polynomial is a multiple of minimal polynomial

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

Proof The desired result follows immediately from the Cayley–Hamilton Theorem (8.37) and 8.46. ■

We know (at least when $\mathbf{F} = \mathbf{C}$) that the zeros of the characteristic polynomial of T are the eigenvalues of T (see 8.36). Now we show that the minimal polynomial has the same zeros (although the multiplicities of these zeros may differ).

8.49 Eigenvalues are the zeros of the minimal polynomial

Let $T \in \mathcal{L}(V)$. Then the zeros of the minimal polynomial of T are precisely the eigenvalues of T.

Proof Let

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

be the minimal polynomial of T.

First suppose $\lambda \in \mathbf{F}$ is a zero of p. Then p can be written in the form

$$p(z) = (z - \lambda)q(z),$$

where q is a monic polynomial with coefficients in $\mathbf{F}$ (see 4.11). Because $p(T) = 0$, we have

$$0 = (T - \lambda I)(q(T)v)$$

for all $v \in V$. Because the degree of q is less than the degree of the minimal polynomial p, there exists at least one vector $v \in V$ such that $q(T)v \neq 0$. The equation above thus implies that λ is an eigenvalue of T, as desired.

To prove the other direction, now suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T. Thus there exists $v \in V$ with $v \neq 0$ such that $Tv = \lambda v$. Repeated applications of T to both sides of this equation show that $T^j v = \lambda^j v$ for every nonnegative integer j. Thus

$$\begin{aligned} 0 = p(T)v &= (a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1}T^{m-1} + T^m)v \\ &= (a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{m-1}\lambda^{m-1} + \lambda^m)v \\ &= p(\lambda)v. \end{aligned}$$

Because $v \neq 0$, the equation above implies that $p(\lambda) = 0$, as desired. ■

The next three examples show how our results can be useful in finding minimal polynomials and in understanding why eigenvalues of some operators cannot be exactly computed.

8.50 **Example** Find the minimal polynomial of the operator $T \in \mathcal{L}(\mathbf{C}^3)$ in Example 8.30.

Solution In Example 8.30 we noted that the eigenvalues of T are 6 and 7. Thus by 8.49, the minimal polynomial of T is a polynomial multiple of $(z - 6)(z - 7)$.

In Example 8.35, we saw that the characteristic polynomial of T is $(z - 6)^2(z - 7)$. Thus by 8.48 and the paragraph above, the minimal polynomial of T is either $(z - 6)(z - 7)$ or $(z - 6)^2(z - 7)$. A simple computation shows that

$$(T - 6I)(T - 7I) \neq 0.$$

Thus the minimal polynomial of T is $(z - 6)^2(z - 7)$.

8.51 Example Find the minimal polynomial of the operator $T \in \mathcal{L}(\mathbf{C}^3)$ defined by $T(z_1, z_2, z_3) = (6z_1, 6z_2, 7z_3)$.

Solution It is easy to see that for this operator T, the eigenvalues of T are 6 and 7, and the characteristic polynomial of T is $(z-6)^2(z-7)$.

Thus as in the previous example, the minimal polynomial of T is either $(z-6)(z-7)$ or $(z-6)^2(z-7)$. A simple computation shows that $(T-6I)(T-7I) = 0$. Thus the minimal polynomial of T is $(z-6)(z-7)$.

8.52 Example What are the eigenvalues of the operator in Example 8.45?

Solution From 8.49 and the solution to Example 8.45, we see that the eigenvalues of T equal the solutions to the equation

$$z^5 - 6z + 3 = 0.$$

Unfortunately, no solution to this equation can be computed using rational numbers, roots of rational numbers, and the usual rules of arithmetic (a proof of this would take us considerably beyond linear algebra). Thus we cannot find an exact expression for any eigenvalue of T in any familiar form, although numeric techniques can give good approximations for the eigenvalues of T. The numeric techniques, which we will not discuss here, show that the eigenvalues for this particular operator are approximately

$$-1.67, \quad 0.51, \quad 1.40, \quad -0.12 + 1.59i, \quad -0.12 - 1.59i.$$

The nonreal eigenvalues occur as a pair, with each the complex conjugate of the other, as expected for a polynomial with real coefficients (see 4.15).

EXERCISES 8.C

1 Suppose $T \in \mathcal{L}(\mathbf{C}^4)$ is such that the eigenvalues of T are 3, 5, 8. Prove that $(T-3I)^2(T-5I)^2(T-8I)^2 = 0$.

2 Suppose V is a complex vector space. Suppose $T \in \mathcal{L}(V)$ is such that 5 and 6 are eigenvalues of T and that T has no other eigenvalues. Prove that $(T-5I)^{n-1}(T-6I)^{n-1} = 0$, where $n = \dim V$.

3 Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $(z-7)^2(z-8)^2$.

4 Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $(z-1)(z-5)^3$ and whose minimal polynomial equals $(z-1)(z-5)^2$.

5 Give an example of an operator on $\mathbf{C}^4$ whose characteristic and minimal polynomials both equal $z(z-1)^2(z-3)$.

6 Give an example of an operator on $\mathbf{C}^4$ whose characteristic polynomial equals $z(z-1)^2(z-3)$ and whose minimal polynomial equals $z(z-1)(z-3)$.

7 Suppose V is a complex vector space. Suppose $T \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that the characteristic polynomial of P is $z^m(z-1)^n$, where $m = \dim \operatorname{null} P$ and $n = \dim \operatorname{range} P$.

8 Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if the constant term in the minimal polynomial of T is nonzero.

9 Suppose $T \in \mathcal{L}(V)$ has minimal polynomial $4+5z-6z^2-7z^3+2z^4+z^5$. Find the minimal polynomial of T^{-1}.

10 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Let p denote the characteristic polynomial of T and let q denote the characteristic polynomial of T^{-1}. Prove that

$$q(z) = \frac{1}{p(0)} z^{\dim V} p\Big(\frac{1}{z}\Big)$$

for all nonzero $z \in \mathbf{C}$.

11 Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that there exists a polynomial $p \in \mathcal{P}(\mathbf{F})$ such that $T^{-1} = p(T)$.

12 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that V has a basis consisting of eigenvectors of T if and only if the minimal polynomial of T has no repeated zeros.
[*For complex vector spaces, the exercise above adds another equivalence to the list given by 5.41.*]

13 Suppose V is an inner product space and $T \in \mathcal{L}(V)$ is normal. Prove that the minimal polynomial of T has no repeated zeros.

14 Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$ is an isometry. Prove that the constant term in the characteristic polynomial of S has absolute value 1.

15 Suppose $T \in \mathcal{L}(V)$ and $v \in V$.

(a) Prove that there exists a unique monic polynomial p of smallest degree such that $p(T)v = 0$.

(b) Prove that p divides the minimal polynomial of T.

16 Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Suppose

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

is the minimal polynomial of T. Prove that

$$\overline{a_0} + \overline{a_1} z + \overline{a_2} z^2 + \cdots + \overline{a_{m-1}} z^{m-1} + z^m$$

is the minimal polynomial of T^*.

17 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Suppose the minimal polynomial of T has degree $\dim V$. Prove that the characteristic polynomial of T equals the minimal polynomial of T.

18 Suppose $a_0, \dots, a_{n-1} \in \mathbf{C}$. Find the minimal and characteristic polynomials of the operator on $\mathbf{C}^n$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & -a_2 \\ & & \ddots & & \vdots \\ & & & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{pmatrix}.$$

[*The exercise above shows that every monic polynomial is the characteristic polynomial of some operator.*]

19 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Suppose that with respect to some basis of V the matrix of T is upper triangular, with $\lambda_1, \dots, \lambda_n$ on the diagonal of this matrix. Prove that the characteristic polynomial of T is $(z - \lambda_1) \cdots (z - \lambda_n)$.

20 Suppose V is a complex vector space and $V_1, \dots, V_m$ are nonzero subspaces of V such that $V = V_1 \oplus \cdots \oplus V_m$. Suppose $T \in \mathcal{L}(V)$ and each V_j is invariant under T. For each j, let p_j denote the characteristic polynomial of $T|_{V_j}$. Prove that the characteristic polynomial of T equals $p_1 \cdots p_m$.

8.D Jordan Form

We know that if V is a complex vector space, then for every $T \in \mathcal{L}(V)$ there is a basis of V with respect to which T has a nice upper-triangular matrix (see 8.29). In this section we will see that we can do even better—there is a basis of V with respect to which the matrix of T contains 0's everywhere except possibly on the diagonal and the line directly above the diagonal.

We begin by looking at two examples of nilpotent operators.

8.53 Example Let $N \in \mathcal{L}(\mathbf{F}^4)$ be the nilpotent operator defined by

$$N(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3).$$

If $v = (1, 0, 0, 0)$, then N^3v, N^2v, Nv, v is a basis of $\mathbf{F}^4$. The matrix of N with respect to this basis is

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The next example of a nilpotent operator has more complicated behavior than the example above.

8.54 Example Let $N \in \mathcal{L}(\mathbf{F}^6)$ be the nilpotent operator defined by

$$N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Unlike the nice behavior of the nilpotent operator of the previous example, for this nilpotent operator there does not exist a vector $v \in \mathbf{F}^6$ such that $N^5v, N^4v, N^3v, N^2v, Nv, v$ is a basis of $\mathbf{F}^6$. However, if we take $v_1 = (1, 0, 0, 0, 0, 0)$, $v_2 = (0, 0, 0, 1, 0, 0)$, and $v_3 = (0, 0, 0, 0, 0, 1)$, then $N^2v_1, Nv_1, v_1, Nv_2, v_2, v_3$ is a basis of $\mathbf{F}^6$. The matrix of N with respect to this basis is

$$\left(\begin{array}{ccc|cc|c} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Here the inner matrices are blocked off to show that we can think of the 6-by-6 matrix above as a block diagonal matrix consisting of a 3-by-3 block with 1's on the line above the diagonal and 0's elsewhere, a 2-by-2 block with 1 above the diagonal and 0's elsewhere, and a 1-by-1 block containing 0.

Our next result shows that every nilpotent operator $N \in \mathcal{L}(V)$ behaves similarly to the previous example. Specifically, there is a finite collection of vectors $v_1, \dots, v_n \in V$ such that there is a basis of V consisting of the vectors of the form $N^k v_j$, as j varies from 1 to n and k varies (in reverse order) from 0 to the largest nonnegative integer m_j such that $N^{m_j} v_j \neq 0$. For the matrix interpretation of the next result, see the first part of the proof of 8.60.

8.55 Basis corresponding to a nilpotent operator

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \dots, v_n \in V$ and nonnegative integers $m_1, \dots, m_n$ such that

(a) $N^{m_1} v_1, \dots, N v_1, v_1, \dots, N^{m_n} v_n, \dots, N v_n, v_n$ is a basis of V;

(b) $N^{m_1+1} v_1 = \cdots = N^{m_n+1} v_n = 0$.

Proof We will prove this result by induction on $\dim V$. To get started, note that the desired result obviously holds if $\dim V = 1$ (in that case, the only nilpotent operator is the 0 operator, so take v_1 to be any nonzero vector and $m_1 = 0$). Now assume that $\dim V > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Because N is nilpotent, N is not injective. Thus N is not surjective (by 3.69) and hence $\operatorname{range} N$ is a subspace of V that has a smaller dimension than V. Thus we can apply our induction hypothesis to the restriction operator $N|_{\operatorname{range} N} \in \mathcal{L}(\operatorname{range} N)$. [We can ignore the trivial case $\operatorname{range} N = \{0\}$, because in that case N is the 0 operator and we can choose $v_1, \dots, v_n$ to be any basis of V and $m_1 = \cdots = m_n = 0$ to get the desired result.]

By our induction hypothesis applied to $N|_{\operatorname{range} N}$, there exist vectors $v_1, \dots, v_n \in \operatorname{range} N$ and nonnegative integers $m_1, \dots, m_n$ such that

8.56 $$N^{m_1} v_1, \dots, N v_1, v_1, \dots, N^{m_n} v_n, \dots, N v_n, v_n$$

is a basis of $\operatorname{range} N$ and

$$N^{m_1+1} v_1 = \cdots = N^{m_n+1} v_n = 0.$$

Because each v_j is in $\operatorname{range} N$, for each j there exists $u_j \in V$ such that $v_j = N u_j$. Thus $N^{k+1} u_j = N^k v_j$ for each j and each nonnegative integer k. We now claim that

8.57 $$N^{m_1+1} u_1, \dots, N u_1, u_1, \dots, N^{m_n+1} u_n, \dots, N u_n, u_n$$

is a linearly independent list of vectors in V. To verify this claim, suppose that some linear combination of 8.57 equals 0. Applying N to that linear combination, we get a linear combination of 8.56 equal to 0. However, the list 8.56 is linearly independent, and hence all the coefficients in our original linear combination of 8.57 equal 0 except possibly the coefficients of the vectors

$$N^{m_1+1}u_1, \dots, N^{m_n+1}u_n,$$

which equal the vectors

$$N^{m_1}v_1, \dots, N^{m_n}v_n.$$

Again using the linear independence of the list 8.56, we conclude that those coefficients also equal 0, completing our proof that the list 8.57 is linearly independent.

Now extend 8.57 to a basis

8.58 $$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, w_1, \dots, w_p$$

of V (which is possible by 2.33). Each Nw_j is in range N and hence is in the span of 8.56. Each vector in the list 8.56 equals N applied to some vector in the list 8.57. Thus there exists x_j in the span of 8.57 such that $Nw_j = Nx_j$. Now let

$$u_{n+j} = w_j - x_j.$$

Then $Nu_{n+j} = 0$. Furthermore,

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, u_{n+1}, \dots, u_{n+p}$$

spans V because its span contains each x_j and each u_{n+j} and hence each w_j (and because 8.58 spans V).

Thus the spanning list above is a basis of V because it has the same length as the basis 8.58 (where we have used 2.42). This basis has the required form, completing the proof. ■

French mathematician Camille Jordan (1838–1922) first published a proof of 8.60 in 1870.

In the next definition, the diagonal of each A_j is filled with some eigenvalue λ_j of T, the line directly above the diagonal of A_j is filled with 1's, and all other entries in A_j are 0 (to understand why each λ_j is an eigenvalue of T, see 5.32). The λ_j's need not be distinct. Also, A_j may be a 1-by-1 matrix (λ_j) containing just an eigenvalue of T.

8.59 Definition ***Jordan basis***

Suppose $T \in \mathcal{L}(V)$. A basis of V is called a ***Jordan basis*** for T if with respect to this basis T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

where each A_j is an upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}.$$

8.60 Jordan Form

Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.

Proof First consider a nilpotent operator $N \in \mathcal{L}(V)$ and the vectors $v_1, \dots, v_n \in V$ given by 8.55. For each j, note that N sends the first vector in the list $N^{m_j} v_j, \dots, N v_j, v_j$ to 0 and that N sends each vector in this list other than the first vector to the previous vector. In other words, 8.55 gives a basis of V with respect to which N has a block diagonal matrix, where each matrix on the diagonal has the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}.$$

Thus the desired result holds for nilpotent operators.

Now suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. We have the generalized eigenspace decomposition

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent (see 8.21). Thus some basis of each $G(\lambda_j, T)$ is a Jordan basis for $(T - \lambda_j I)|_{G(\lambda_m, T)}$ (see previous paragraph). Put these bases together to get a basis of V that is a Jordan basis for T. ■

EXERCISES 8.D

1 Find the characteristic polynomial and the minimal polynomial of the operator N in Example 8.53.

2 Find the characteristic polynomial and the minimal polynomial of the operator N in Example 8.54.

3 Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that the minimal polynomial of N is z^{m+1}, where m is the length of the longest consecutive string of 1's that appears on the line directly above the diagonal in the matrix of N with respect to any Jordan basis for N.

4 Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V that is a Jordan basis for T. Describe the matrix of T with respect to the basis $v_n, \ldots, v_1$ obtained by reversing the order of the v's.

5 Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V that is a Jordan basis for T. Describe the matrix of T^2 with respect to this basis.

6 Suppose $N \in \mathcal{L}(V)$ is nilpotent and $v_1, \ldots, v_n$ and $m_1, \ldots, m_n$ are as in 8.55. Prove that $N^{m_1} v_1, \ldots, N^{m_n} v_n$ is a basis of null N.
[*The exercise above implies that n, which equals* $\dim \operatorname{null} N$, *depends only on N and not on the specific Jordan basis chosen for N.*]

7 Suppose $p, q \in \mathcal{P}(\mathbf{C})$ are monic polynomials with the same zeros and q is a polynomial multiple of p. Prove that there exists $T \in \mathcal{L}(\mathbf{C}^{\deg q})$ such that the characteristic polynomial of T is q and the minimal polynomial of T is p.

8 Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Prove that there does not exist a direct sum decomposition of V into two proper subspaces invariant under T if and only if the minimal polynomial of T is of the form $(z - \lambda)^{\dim V}$ for some $\lambda \in \mathbf{C}$.

CHAPTER

9

Euclid explaining geometry (from The School of Athens, *painted by Raphael around 1510).*

Operators on Real Vector Spaces

In the last chapter we learned about the structure of an operator on a finite-dimensional complex vector space. In this chapter, we will use our results about operators on complex vector spaces to learn about operators on real vector spaces.

Our assumptions for this chapter are as follows:

9.1 Notation F, V

- **F** denotes **R** or **C**.
- V denotes a finite-dimensional nonzero vector space over **F**.

LEARNING OBJECTIVES FOR THIS CHAPTER

- complexification of a real vector space
- complexification of an operator on a real vector space
- operators on finite-dimensional real vector spaces have an eigenvalue or a 2-dimensional invariant subspace
- characteristic polynomial and the Cayley–Hamilton Theorem
- description of normal operators on a real inner product space
- description of isometries on a real inner product space

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_9

9.A Complexification

Complexification of a Vector Space

As we will soon see, a real vector space V can be embedded, in a natural way, in a complex vector space called the complexification of V. Each operator on V can be extended to an operator on the complexification of V. Our results about operators on complex vector spaces can then be translated to information about operators on real vector spaces.

We begin by defining the complexification of a real vector space.

9.2 Definition ***complexification of*** V**,** $V_{\mathbf{C}}$

Suppose V is a real vector space.

- The ***complexification*** of V, denoted $V_{\mathbf{C}}$, equals $V \times V$. An element of $V_{\mathbf{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we will write this as $u + iv$.
- Addition on $V_{\mathbf{C}}$ is defined by
$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$
for $u_1, v_1, u_2, v_2 \in V$.
- Complex scalar multiplication on $V_{\mathbf{C}}$ is defined by
$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$
for $a, b \in \mathbf{R}$ and $u, v \in V$.

Motivation for the definition above of complex scalar multiplication comes from usual algebraic properties and the identity $i^2 = -1$. If you remember the motivation, then you do not need to memorize the definition above.

We think of V as a subset of $V_{\mathbf{C}}$ by identifying $u \in V$ with $u + i0$. The construction of $V_{\mathbf{C}}$ from V can then be thought of as generalizing the construction of $\mathbf{C}^n$ from $\mathbf{R}^n$.

9.3 $V_{\mathbf{C}}$ **is a complex vector space.**

Suppose V is a real vector space. Then with the definitions of addition and scalar multiplication as above, $V_{\mathbf{C}}$ is a complex vector space.

The proof of the result above is left as an exercise for the reader. Note that the additive identity of $V_{\mathbf{C}}$ is $0 + i0$, which we write as just 0.

Probably everything that you think should work concerning complexification does work, usually with a straightforward verification, as illustrated by the next result.

9.4 Basis of V is basis of $V_{\mathbf{C}}$

Suppose V is a real vector space.

(a) If $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is a basis of $V_{\mathbf{C}}$ (as a complex vector space).

(b) The dimension of $V_{\mathbf{C}}$ (as a complex vector space) equals the dimension of V (as a real vector space).

Proof To prove (a), suppose $v_1, \ldots, v_n$ is a basis of the real vector space V. Then $\operatorname{span}(v_1, \ldots, v_n)$ in the complex vector space $V_{\mathbf{C}}$ contains all the vectors $v_1, \ldots, v_n, iv_1, \ldots, iv_n$. Thus $v_1, \ldots, v_n$ spans the complex vector space $V_{\mathbf{C}}$.

To show that $v_1, \ldots, v_n$ is linearly independent in the complex vector space $V_{\mathbf{C}}$, suppose $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$ and

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0.$$

Then the equation above and our definitions imply that

$$(\operatorname{Re}\lambda_1)v_1 + \cdots + (\operatorname{Re}\lambda_n)v_n = 0 \quad \text{and} \quad (\operatorname{Im}\lambda_1)v_1 + \cdots + (\operatorname{Im}\lambda_n)v_n = 0.$$

Because $v_1, \ldots, v_n$ is linearly independent in V, the equations above imply $\operatorname{Re}\lambda_1 = \cdots = \operatorname{Re}\lambda_n = 0$ and $\operatorname{Im}\lambda_1 = \cdots = \operatorname{Im}_n = 0$. Thus we have $\lambda_1 = \cdots = \lambda_n = 0$. Hence $v_1, \ldots, v_n$ is linearly independent in $V_{\mathbf{C}}$, completing the proof of (a).

Clearly (b) follows immediately from (a). ■

Complexification of an Operator

Now we can define the complexification of an operator.

9.5 Definition ***complexification of T, $T_{\mathbf{C}}$***

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. The ***complexification*** of T, denoted $T_{\mathbf{C}}$, is the operator $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ defined by

$$T_{\mathbf{C}}(u + iv) = Tu + iTv$$

for $u, v \in V$.

You should verify that if V is a real vector space and $T \in \mathcal{L}(V)$, then $T_{\mathbf{C}}$ is indeed in $\mathcal{L}(V_{\mathbf{C}})$. The key point here is that our definition of complex scalar multiplication can be used to show that $T_{\mathbf{C}}\big(\lambda(u + iv)\big) = \lambda T_{\mathbf{C}}(u + iv)$ for all $u, v \in V$ and all **complex** numbers λ.

The next example gives a good way to think about the complexification of a typical operator.

9.6 **Example** Suppose A is an n-by-n matrix of real numbers. Define $T \in \mathcal{L}(\mathbf{R}^n)$ by $Tx = Ax$, where elements of $\mathbf{R}^n$ are thought of as n-by-1 column vectors. Identifying the complexification of $\mathbf{R}^n$ with $\mathbf{C}^n$, we then have $T_{\mathbf{C}}z = Az$ for each $z \in \mathbf{C}^n$, where again elements of $\mathbf{C}^n$ are thought of as n-by-1 column vectors.

In other words, if T is the operator of matrix multiplication by A on $\mathbf{R}^n$, then the complexification $T_{\mathbf{C}}$ is also matrix multiplication by A but now acting on the larger domain $\mathbf{C}^n$.

The next result makes sense because 9.4 tells us that a basis of a real vector space is also a basis of its complexification. The proof of the next result follows immediately from the definitions.

9.7 Matrix of $T_{\mathbf{C}}$ equals matrix of T

Suppose V is a real vector space with basis $v_1, \dots, v_n$ and $T \in \mathcal{L}(V)$. Then $\mathcal{M}(T) = \mathcal{M}(T_{\mathbf{C}})$, where both matrices are with respect to the basis $v_1, \dots, v_n$.

The result above and Example 9.6 provide complete insight into complexification, because once a basis is chosen, every operator essentially looks like Example 9.6. Complexification of an operator could have been defined using matrices, but the approach taken here is more natural because it does not depend on the choice of a basis.

We know that every operator on a nonzero finite-dimensional complex vector space has an eigenvalue (see 5.21) and thus has a 1-dimensional invariant subspace. We have seen an example [5.8(a)] of an operator on a nonzero finite-dimensional real vector space with no eigenvalues and thus no 1-dimensional invariant subspaces. However, we now show that an invariant subspace of dimension 1 or 2 always exists. Notice how complexification leads to a simple proof of this result.

9.8 Every operator has an invariant subspace of dimension 1 or 2

Every operator on a nonzero finite-dimensional vector space has an invariant subspace of dimension 1 or 2.

Proof Every operator on a nonzero finite-dimensional complex vector space has an eigenvalue (5.21) and thus has a 1-dimensional invariant subspace.

Hence assume V is a real vector space and $T \in \mathcal{L}(V)$. The complexification $T_{\mathbf{C}}$ has an eigenvalue $a + bi$ (by 5.21), where $a, b \in \mathbf{R}$. Thus there exist $u, v \in V$, not both 0, such that $T_{\mathbf{C}}(u + iv) = (a + bi)(u + iv)$. Using the definition of $T_{\mathbf{C}}$, the last equation can be rewritten as

$$Tu + iTv = (au - bv) + (av + bu)i.$$

Thus

$$Tu = au - bv \quad \text{and} \quad Tv = av + bu.$$

Let U equal the span in V of the list u, v. Then U is a subspace of V with dimension 1 or 2. The equations above show that U is invariant under T, completing the proof. ■

The Minimal Polynomial of the Complexification

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Repeated application of the definition of $T_{\mathbf{C}}$ shows that

9.9 $$(T_{\mathbf{C}})^n(u + iv) = T^n u + iT^n v$$

for every positive integer n and all $u, v \in V$.

Notice that the next result implies that the minimal polynomial of $T_{\mathbf{C}}$ has real coefficients.

9.10 Minimal polynomial of $T_{\mathbf{C}}$ equals minimal polynomial of T

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the minimal polynomial of $T_{\mathbf{C}}$ equals the minimal polynomial of T.

Proof Let $p \in \mathcal{P}(\mathbf{R})$ denote the minimal polynomial of T. From 9.9 it is easy to see that $p(T_{\mathbf{C}}) = \big(p(T)\big)_{\mathbf{C}}$, and thus $p(T_{\mathbf{C}}) = 0$.

Suppose $q \in \mathcal{P}(\mathbf{C})$ is a monic polynomial such that $q(T_{\mathbf{C}}) = 0$. Then $\big(q(T_{\mathbf{C}})\big)(u) = 0$ for every $u \in V$. Letting r denote the polynomial whose j^{th} coefficient is the real part of the j^{th} coefficient of q, we see that r is a monic polynomial and $r(T) = 0$. Thus $\deg q = \deg r \geq \deg p$.

The conclusions of the two previous paragraphs imply that p is the minimal polynomial of $T_{\mathbf{C}}$, as desired. ■

Eigenvalues of the Complexification

Now we turn to questions about the eigenvalues of the complexification of an operator. Again, everything that we expect to work indeed works easily.

We begin with a result showing that the real eigenvalues of $T_{\mathbf{C}}$ are precisely the eigenvalues of T. We give two different proofs of this result. The first proof is more elementary, but the second proof is shorter and gives some useful insight.

9.11 Real eigenvalues of $T_{\mathbf{C}}$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{R}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if λ is an eigenvalue of T.

Proof 1 First suppose λ is an eigenvalue of T. Then there exists $v \in V$ with $v \neq 0$ such that $Tv = \lambda v$. Thus $T_{\mathbf{C}}v = \lambda v$, which shows that λ is an eigenvalue of $T_{\mathbf{C}}$, completing one direction of the proof.

To prove the other direction, suppose now that λ is an eigenvalue of $T_{\mathbf{C}}$. Then there exist $u, v \in V$ with $u + iv \neq 0$ such that

$$T_{\mathbf{C}}(u + iv) = \lambda(u + iv).$$

The equation above implies that $Tu = \lambda u$ and $Tv = \lambda v$. Because $u \neq 0$ or $v \neq 0$, this implies that λ is an eigenvalue of T, completing the proof. ■

Proof 2 The (real) eigenvalues of T are the (real) zeros of the minimal polynomial of T (by 8.49). The real eigenvalues of $T_{\mathbf{C}}$ are the real zeros of the minimal polynomial of $T_{\mathbf{C}}$ (again by 8.49). These two minimal polynomials are the same (by 9.10). Thus the eigenvalues of T are precisely the real eigenvalues of $T_{\mathbf{C}}$, as desired. ■

Our next result shows that $T_{\mathbf{C}}$ behaves symmetrically with respect to an eigenvalue λ and its complex conjugate $\bar{\lambda}$.

9.12 $T_{\mathbf{C}} - \lambda I$ and $T_{\mathbf{C}} - \bar{\lambda} I$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{C}$, j is a nonnegative integer, and $u, v \in V$. Then

$$(T_{\mathbf{C}} - \lambda I)^j(u + iv) = 0 \quad \text{if and only if} \quad (T_{\mathbf{C}} - \bar{\lambda} I)^j(u - iv) = 0.$$

Proof We will prove this result by induction on j. To get started, note that if $j = 0$ then (because an operator raised to the power 0 equals the identity operator) the result claims that $u + iv = 0$ if and only if $u - iv = 0$, which is clearly true.

Thus assume by induction that $j \geq 1$ and the desired result holds for $j - 1$. Suppose $(T_{\mathbf{C}} - \lambda I)^j(u + iv) = 0$. Then

9.13 $$(T_{\mathbf{C}} - \lambda I)^{j-1}\big((T_{\mathbf{C}} - \lambda I)(u + iv)\big) = 0.$$

Writing $\lambda = a + bi$, where $a, b \in \mathbf{R}$, we have

9.14 $$(T_{\mathbf{C}} - \lambda I)(u + iv) = (Tu - au + bv) + i(Tv - av - bu)$$

and

9.15 $$(T_{\mathbf{C}} - \bar{\lambda} I)(u - iv) = (Tu - au + bv) - i(Tv - av - bu).$$

Our induction hypothesis, 9.13, and 9.14 imply that

$$(T_{\mathbf{C}} - \bar{\lambda} I)^{j-1}\big((Tu - au + bv) - i(Tv - av - bu)\big) = 0.$$

Now the equation above and 9.15 imply that $(T_{\mathbf{C}} - \bar{\lambda} I)^j(u - iv) = 0$, completing the proof in one direction.

The other direction is proved by replacing λ with $\bar{\lambda}$, replacing v with $-v$, and then using the first direction. ■

An important consequence of the result above is the next result, which states that if a number is an eigenvalue of $T_{\mathbf{C}}$, then its complex conjugate is also an eigenvalue of $T_{\mathbf{C}}$.

9.16 Nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if $\bar{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$.

Proof Take $j = 1$ in 9.12. ■

By definition, the eigenvalues of an operator on a real vector space are real numbers. Thus when mathematicians sometimes informally mention the complex eigenvalues of an operator on a real vector space, what they have in mind is the eigenvalues of the complexification of the operator.

Recall that the multiplicity of an eigenvalue is defined to be the dimension of the generalized eigenspace corresponding to that eigenvalue (see 8.24). The next result states that the multiplicity of an eigenvalue of a complexification equals the multiplicity of its complex conjugate.

9.17 Multiplicity of λ equals multiplicity of $\bar{\lambda}$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$ is an eigenvalue of $T_{\mathbf{C}}$. Then the multiplicity of λ as an eigenvalue of $T_{\mathbf{C}}$ equals the multiplicity of $\bar{\lambda}$ as an eigenvalue of $T_{\mathbf{C}}$.

Proof Suppose $u_1 + iv_1, \dots, u_m + iv_m$ is a basis of the generalized eigenspace $G(\lambda, T_{\mathbf{C}})$, where $u_1, \dots, u_m, v_1, \dots, v_m \in V$. Then using 9.12, routine arguments show that $u_1 - iv_1, \dots, u_m - iv_m$ is a basis of the generalized eigenspace $G(\bar{\lambda}, T_{\mathbf{C}})$. Thus both λ and $\bar{\lambda}$ have multiplicity m as eigenvalues of $T_{\mathbf{C}}$. ■

9.18 Example Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ is defined by

$$T(x_1, x_2, x_3) = (2x_1, x_2 - x_3, x_2 + x_3).$$

The matrix of T with respect to the standard basis of $\mathbf{R}^3$ is $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$.

As you can verify, 2 is an eigenvalue of T with multiplicity 1 and T has no other eigenvalues.

If we identify the complexification of $\mathbf{R}^3$ with $\mathbf{C}^3$, then the matrix of $T_{\mathbf{C}}$ with respect to the standard basis of $\mathbf{C}^3$ is the matrix above. As you can verify, the eigenvalues of $T_{\mathbf{C}}$ are 2, $1 + i$, and $1 - i$, each with multiplicity 1. Thus the nonreal eigenvalues of $T_{\mathbf{C}}$ come as a pair, with each the complex conjugate of the other and with the same multiplicity, as expected by 9.17.

We have seen an example [5.8(a)] of an operator on $\mathbf{R}^2$ with no eigenvalues. The next result shows that no such example exists on $\mathbf{R}^3$.

9.19 Operator on odd-dimensional vector space has eigenvalue

Every operator on an odd-dimensional real vector space has an eigenvalue.

Proof Suppose V is a real vector space with odd dimension and $T \in \mathcal{L}(V)$. Because the nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs with equal multiplicity (by 9.17), the sum of the multiplicities of all the nonreal eigenvalues of $T_{\mathbf{C}}$ is an even number.

Because the sum of the multiplicities of all the eigenvalues of $T_{\mathbf{C}}$ equals the (complex) dimension of $V_{\mathbf{C}}$ (by Theorem 8.26), the conclusion of the paragraph above implies that $T_{\mathbf{C}}$ has a real eigenvalue. Every real eigenvalue of $T_{\mathbf{C}}$ is also an eigenvalue of T (by 9.11), giving the desired result. ■

Characteristic Polynomial of the Complexification

In the previous chapter we defined the characteristic polynomial of an operator on a finite-dimensional complex vector space (see 8.34). The next result is a key step toward defining the characteristic polynomial for operators on finite-dimensional real vector spaces.

9.20 Characteristic polynomial of $T_{\mathbf{C}}$

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real.

Proof Suppose λ is a nonreal eigenvalue of $T_{\mathbf{C}}$ with multiplicity m. Then $\bar{\lambda}$ is also an eigenvalue of $T_{\mathbf{C}}$ with multiplicity m (by 9.17). Thus the characteristic polynomial of $T_{\mathbf{C}}$ includes factors of $(z-\lambda)^m$ and $(z-\bar{\lambda})^m$. Multiplying together these two factors, we have

$$(z-\lambda)^m(z-\bar{\lambda})^m = \big(z^2 - 2(\operatorname{Re}\lambda)z + |\lambda|^2\big)^m.$$

The polynomial above on the right has real coefficients.

The characteristic polynomial of $T_{\mathbf{C}}$ is the product of terms of the form above and terms of the form $(z-t)^d$, where t is a real eigenvalue of $T_{\mathbf{C}}$ with multiplicity d. Thus the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real. ■

Now we can define the characteristic polynomial of an operator on a finite-dimensional real vector space to be the characteristic polynomial of its complexification.

9.21 Definition ***Characteristic polynomial***

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the ***characteristic polynomial*** of T is defined to be the characteristic polynomial of $T_{\mathbf{C}}$.

9.22 Example Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ is defined by

$$T(x_1, x_2, x_3) = (2x_1, x_2 - x_3, x_2 + x_3).$$

As we noted in 9.18, the eigenvalues of $T_{\mathbf{C}}$ are 2, $1+i$, and $1-i$, each with multiplicity 1. Thus the characteristic polynomial of the complexification $T_{\mathbf{C}}$ is $(z-2)\big(z-(1+i)\big)\big(z-(1-i)\big)$, which equals $z^3 - 4z^2 + 6z - 4$. Hence the characteristic polynomial of T is also $z^3 - 4z^2 + 6z - 4$.

In the next result, the eigenvalues of T are all real (because T is an operator on a real vector space).

9.23 Degree and zeros of characteristic polynomial

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then

(a) the coefficients of the characteristic polynomial of T are all real;

(b) the characteristic polynomial of T has degree $\dim V$;

(c) the eigenvalues of T are precisely the real zeros of the characteristic polynomial of T.

Proof Part (a) holds because of 9.20.

Part (b) follows from 8.36(a).

Part (c) holds because the real zeros of the characteristic polynomial of T are the real eigenvalues of $T_{\mathbf{C}}$ [by 8.36(a)], which are the eigenvalues of T (by 9.11). ■

In the previous chapter, we proved the Cayley–Hamilton Theorem (8.37) for complex vector spaces. Now we can also prove it for real vector spaces.

9.24 Cayley–Hamilton Theorem

Suppose $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

Proof We have already proved this result when V is a complex vector space. Thus assume that V is a real vector space.

The complex case of the Cayley–Hamilton Theorem (8.37) implies that $q(T_{\mathbf{C}}) = 0$. Thus we also have $q(T) = 0$, as desired. ■

9.25 **Example** Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ is defined by

$$T(x_1, x_2, x_3) = (2x_1, x_2 - x_3, x_2 + x_3).$$

As we saw in 9.22, the characteristic polynomial of T is $z^3 - 4z^2 + 6z - 4$. Thus the Cayley–Hamilton Theorem implies that $T^3 - 4T^2 + 6T - 4I = 0$, which can also be verified by direct calculation.

We can now prove another result that we previously knew only in the complex case.

9.26 Characteristic polynomial is a multiple of minimal polynomial

Suppose $T \in \mathcal{L}(V)$. Then

(a) the degree of the minimal polynomial of T is at most $\dim V$;

(b) the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

Proof Part (a) follows immediately from the Cayley–Hamilton Theorem. Part (b) follows from the Cayley–Hamilton Theorem and 8.46. ■

EXERCISES 9.A

1 Prove 9.3.

2 Verify that if V is a real vector space and $T \in \mathcal{L}(V)$, then $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$.

3 Suppose V is a real vector space and $v_1, \dots, v_m \in V$. Prove that $v_1, \dots, v_m$ is linearly independent in $V_{\mathbf{C}}$ if and only if $v_1, \dots, v_m$ is linearly independent in V.

4 Suppose V is a real vector space and $v_1, \dots, v_m \in V$. Prove that $v_1, \dots, v_m$ spans $V_{\mathbf{C}}$ if and only if $v_1, \dots, v_m$ spans V.

5 Suppose that V is a real vector space and $S, T \in \mathcal{L}(V)$. Show that $(S + T)_{\mathbf{C}} = S_{\mathbf{C}} + T_{\mathbf{C}}$ and that $(\lambda T)_{\mathbf{C}} = \lambda T_{\mathbf{C}}$ for every $\lambda \in \mathbf{R}$.

6 Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Prove that $T_{\mathbf{C}}$ is invertible if and only if T is invertible.

7 Suppose V is a real vector space and $N \in \mathcal{L}(V)$. Prove that $N_{\mathbf{C}}$ is nilpotent if and only if N is nilpotent.

8 Suppose $T \in \mathcal{L}(\mathbf{R}^3)$ and $5, 7$ are eigenvalues of T. Prove that $T_{\mathbf{C}}$ has no nonreal eigenvalues.

9 Prove there does not exist an operator $T \in \mathcal{L}(\mathbf{R}^7)$ such that $T^2 + T + I$ is nilpotent.

10 Give an example of an operator $T \in \mathcal{L}(\mathbf{C}^7)$ such that $T^2 + T + I$ is nilpotent.

11 Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Suppose there exist $b, c \in \mathbf{R}$ such that $T^2 + bT + cI = 0$. Prove that T has an eigenvalue if and only if $b^2 \geq 4c$.

12 Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Suppose there exist $b, c \in \mathbf{R}$ such that $b^2 < 4c$ and $T^2 + bT + cI$ is nilpotent. Prove that T has no eigenvalues.

13 Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $b, c \in \mathbf{R}$ are such that $b^2 < 4c$. Prove that $\operatorname{null}(T^2 + bT + cI)^j$ has even dimension for every positive integer j.

14 Suppose V is a real vector space with $\dim V = 8$. Suppose $T \in \mathcal{L}(V)$ is such that $T^2 + T + I$ is nilpotent. Prove that $(T^2 + T + I)^4 = 0$.

15 Suppose V is a real vector space and $T \in \mathcal{L}(V)$ has no eigenvalues. Prove that every subspace of V invariant under T has even dimension.

16 Suppose V is a real vector space. Prove that there exists $T \in \mathcal{L}(V)$ such that $T^2 = -I$ if and only if V has even dimension.

17 Suppose V is a real vector space and $T \in \mathcal{L}(V)$ satisfies $T^2 = -I$. Define complex scalar multiplication on V as follows: if $a, b \in \mathbf{R}$, then

$$(a + bi)v = av + bTv.$$

(a) Show that the complex scalar multiplication on V defined above and the addition on V makes V into a complex vector space.

(b) Show that the dimension of V as a complex vector space is half the dimension of V as a real vector space.

18 Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Prove that the following are equivalent:

(a) All the eigenvalues of $T_{\mathbf{C}}$ are real.

(b) There exists a basis of V with respect to which T has an upper-triangular matrix.

(c) There exists a basis of V consisting of generalized eigenvectors of T.

19 Suppose V is a real vector space with $\dim V = n$ and $T \in \mathcal{L}(V)$ is such that $\operatorname{null} T^{n-2} \neq \operatorname{null} T^{n-1}$. Prove that T has at most two distinct eigenvalues and that $T_{\mathbf{C}}$ has no nonreal eigenvalues.

9.B *Operators on Real Inner Product Spaces*

We now switch our focus to the context of inner product spaces. We will give a complete description of normal operators on real inner product spaces; a key step in the proof of this result (9.34) requires the result from the previous section that an operator on a finite-dimensional real vector space has an invariant subspace of dimension 1 or 2 (9.8).

After describing the normal operators on real inner product spaces, we will use that result to give a complete description of isometries on such spaces.

Normal Operators on Real Inner Product Spaces

The Complex Spectral Theorem (7.24) gives a complete description of normal operators on complex inner product spaces. In this subsection we will give a complete description of normal operators on real inner product spaces.

We begin with a description of the operators on 2-dimensional real inner product spaces that are normal but not self-adjoint.

9.27 Normal but not self-adjoint operators

Suppose V is a 2-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal but not self-adjoint.

(b) The matrix of T with respect to every orthonormal basis of V has the form
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$
with $b \neq 0$.

(c) The matrix of T with respect to some orthonormal basis of V has the form
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$
with $b > 0$.

Proof First suppose (a) holds, so that T is normal but not self-adjoint. Let e_1, e_2 be an orthonormal basis of V. Suppose

9.28 $$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Then $\|Te_1\|^2 = a^2 + b^2$ and $\|T^*e_1\|^2 = a^2 + c^2$. Because T is normal, $\|Te_1\| = \|T^*e_1\|$ (see 7.20); thus these equations imply that $b^2 = c^2$. Thus $c = b$ or $c = -b$. But $c \neq b$, because otherwise T would be self-adjoint, as can be seen from the matrix in 9.28. Hence $c = -b$, so

9.29
$$\mathcal{M}\big(T, (e_1, e_2)\big) = \begin{pmatrix} a & -b \\ b & d \end{pmatrix}.$$

The matrix of T^* is the transpose of the matrix above. Use matrix multiplication to compute the matrices of TT^* and T^*T (do it now). Because T is normal, these two matrices are equal. Equating the entries in the upper-right corner of the two matrices you computed, you will discover that $bd = ab$. Now $b \neq 0$, because otherwise T would be self-adjoint, as can be seen from the matrix in 9.29. Thus $d = a$, completing the proof that (a) implies (b).

Now suppose (b) holds. We want to prove that (c) holds. Choose an orthonormal basis e_1, e_2 of V. We know that the matrix of T with respect to this basis has the form given by (b), with $b \neq 0$. If $b > 0$, then (c) holds and we have proved that (b) implies (c). If $b < 0$, then, as you should verify, the matrix of T with respect to the orthonormal basis $e_1, -e_2$ equals $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$, where $-b > 0$; thus in this case we also see that (b) implies (c).

Now suppose (c) holds, so that the matrix of T with respect to some orthonormal basis has the form given in (c) with $b > 0$. Clearly the matrix of T is not equal to its transpose (because $b \neq 0$). Hence T is not self-adjoint. Now use matrix multiplication to verify that the matrices of TT^* and T^*T are equal. We conclude that $TT^* = T^*T$. Hence T is normal. Thus (c) implies (a), completing the proof. ■

The next result tells us that a normal operator restricted to an invariant subspace is normal. This will allow us to use induction on $\dim V$ when we prove our description of normal operators (9.34).

9.30 Normal operators and invariant subspaces

Suppose V is an inner product space, $T \in \mathcal{L}(V)$ is normal, and U is a subspace of V that is invariant under T. Then

(a) $U^\perp$ is invariant under T;

(b) U is invariant under T^*;

(c) $(T|_U)^* = (T^*)|_U$;

(d) $T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are normal operators.

Proof First we will prove (a). Let $e_1, \dots, e_m$ be an orthonormal basis of U. Extend to an orthonormal basis $e_1, \dots, e_m, f_1, \dots, f_n$ of V (this is possible by 6.35). Because U is invariant under T, each Te_j is a linear combination of $e_1, \dots, e_m$. Thus the matrix of T with respect to the basis $e_1, \dots, e_m, f_1, \dots, f_n$ is of the form

$$\mathcal{M}(T) = \begin{array}{c} \\ e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_n \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & \dots & e_m & f_1 & \dots & f_n \end{array} \\ \left(\begin{array}{cc} A & B \\ & \\ 0 & C \end{array} \right) \end{array};$$

here A denotes an m-by-m matrix, 0 denotes the n-by-m matrix of all 0's, B denotes an m-by-n matrix, C denotes an n-by-n matrix, and for convenience the basis has been listed along the top and left sides of the matrix.

For each $j \in \{1, \dots, m\}$, $\|Te_j\|^2$ equals the sum of the squares of the absolute values of the entries in the j^{th} column of A (see 6.25). Hence

9.31 $$\sum_{j=1}^{m} \|Te_j\|^2 = \begin{array}{l}\text{the sum of the squares of the absolute} \\ \text{values of the entries of } A.\end{array}$$

For each $j \in \{1, \dots, m\}$, $\|T^* e_j\|^2$ equals the sum of the squares of the absolute values of the entries in the j^{th} rows of A and B. Hence

9.32 $$\sum_{j=1}^{m} \|T^* e_j\|^2 = \begin{array}{l}\text{the sum of the squares of the absolute} \\ \text{values of the entries of } A \text{ and } B.\end{array}$$

Because T is normal, $\|Te_j\| = \|T^* e_j\|$ for each j (see 7.20); thus

$$\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m} \|T^* e_j\|^2.$$

This equation, along with 9.31 and 9.32, implies that the sum of the squares of the absolute values of the entries of B equals 0. In other words, B is the matrix of all 0's. Thus

9.33
$$\mathcal{M}(T) = \begin{array}{c} \\ e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_n \end{array} \!\!\!\! \begin{array}{c} \begin{array}{cccccc} e_1 & \dots & e_m & f_1 & \dots & f_n \end{array} \\ \left(\begin{array}{cc} A & 0 \\ & \\ & \\ 0 & C \end{array} \right) \end{array}.$$

This representation shows that Tf_k is in the span of $f_1, \dots, f_n$ for each k. Because $f_1, \dots, f_n$ is a basis of $U^\perp$, this implies that $Tv \in U^\perp$ whenever $v \in U^\perp$. In other words, $U^\perp$ is invariant under T, completing the proof of (a).

To prove (b), note that $\mathcal{M}(T^*)$, which is the conjugate transpose of $\mathcal{M}(T)$, has a block of 0's in the lower left corner (because $\mathcal{M}(T)$, as given above, has a block of 0's in the upper right corner). In other words, each T^*e_j can be written as a linear combination of $e_1, \dots, e_m$. Thus U is invariant under T^*, completing the proof of (b).

To prove (c), let $S = T|_U \in \mathcal{L}(U)$. Fix $v \in U$. Then

$$\begin{aligned} \langle Su, v\rangle &= \langle Tu, v\rangle \\ &= \langle u, T^*v\rangle \end{aligned}$$

for all $u \in U$. Because $T^*v \in U$ [by (b)], the equation above shows that $S^*v = T^*v$. In other words, $(T|_U)^* = (T^*)|_U$, completing the proof of (c).

To prove (d), note that T commutes with T^* (because T is normal) and that $(T|_U)^* = (T^*)|_U$ [by (c)]. Thus $T|_U$ commutes with its adjoint and hence is normal. Interchanging the roles of U and $U^\perp$, which is justified by (a), shows that $T|_{U^\perp}$ is also normal, completing the proof of (d). ■

Note that if an operator T has a block diagonal matrix with respect to some basis, then the entry in each 1-by-1 block on the diagonal of this matrix is an eigenvalue of T.

Our next result shows that normal operators on real inner product spaces come close to having diagonal matrices. Specifically, we get block diagonal matrices, with each block having size at most 2-by-2.

We cannot expect to do better than the next result, because on a real inner product space there exist normal operators that do not have a diagonal matrix with respect to any basis. For example, the operator $T \in \mathcal{L}(\mathbf{R}^2)$ defined by $T(x, y) = (-y, x)$ is normal (as you should verify) but has no eigenvalues; thus this particular T does not have even an upper-triangular matrix with respect to any basis of $\mathbf{R}^2$.

9.34 Characterization of normal operators when $\mathbf{F} = \mathbf{R}$

Suppose V is a real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal.

(b) There is an orthonormal basis of V with respect to which T has a block diagonal matrix such that each block is a 1-by-1 matrix or a 2-by-2 matrix of the form
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$
with $b > 0$.

Proof First suppose (b) holds. With respect to the basis given by (b), the matrix of T commutes with the matrix of T^* (which is the transpose of the matrix of T), as you should verify (use Exercise 9 in Section 8.B for the product of two block diagonal matrices). Thus T commutes with T^*, which means that T is normal, completing the proof that (b) implies (a).

Now suppose (a) holds, so T is normal. We will prove that (b) holds by induction on $\dim V$. To get started, note that our desired result holds if $\dim V = 1$ (trivially) or if $\dim V = 2$ [if T is self-adjoint, use the Real Spectral Theorem (7.29); if T is not self-adjoint, use 9.27].

Now assume that $\dim V > 2$ and that the desired result holds on vector spaces of smaller dimension. Let U be a subspace of V of dimension 1 that is invariant under T if such a subspace exists (in other words, if T has an eigenvector, let U be the span of this eigenvector). If no such subspace exists, let U be a subspace of V of dimension 2 that is invariant under T (an invariant subspace of dimension 1 or 2 always exists by 9.8).

If $\dim U = 1$, choose a vector in U with norm 1; this vector will be an orthonormal basis of U, and of course the matrix of $T|_U \in \mathcal{L}(U)$ is a 1-by-1 matrix. If $\dim U = 2$, then $T|_U \in \mathcal{L}(U)$ is normal (by 9.30) but not self-adjoint (otherwise $T|_U$, and hence T, would have an eigenvector by 7.27). Thus we can choose an orthonormal basis of U with respect to which the matrix of $T|_U \in \mathcal{L}(U)$ has the required form (see 9.27).

Now $U^\perp$ is invariant under T and $T|_{U^\perp}$ is a normal operator on $U^\perp$ (by 9.30). Thus by our induction hypothesis, there is an orthonormal basis of $U^\perp$ with respect to which the matrix of $T|_{U^\perp}$ has the desired form. Adjoining this basis to the basis of U gives an orthonormal basis of V with respect to which the matrix of T has the desired form. Thus (b) holds. ∎

Isometries on Real Inner Product Spaces

As we will see, the next example is a key building block for isometries on real inner product spaces. Also, note that the next example shows that an isometry on $\mathbf{R}^2$ may have no eigenvalues.

9.35 **Example** Let $\theta \in \mathbf{R}$. Then the operator on $\mathbf{R}^2$ of counterclockwise rotation (centered at the origin) by an angle of θ is an isometry, as is geometrically obvious. The matrix of this operator with respect to the standard basis is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

If θ is not an integer multiple of π, then no nonzero vector of $\mathbf{R}^2$ gets mapped to a scalar multiple of itself, and hence the operator has no eigenvalues.

The next result shows that every isometry on a real inner product space is composed of pieces that are rotations on 2-dimensional subspaces, pieces that equal the identity operator, and pieces that equal multiplication by -1.

9.36 **Description of isometries when $\mathbf{F} = \mathbf{R}$**

Suppose V is a real inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry.

(b) There is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block on the diagonal is a 1-by-1 matrix containing 1 or -1 or is a 2-by-2 matrix of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

with $\theta \in (0, \pi)$.

Proof First suppose (a) holds, so S is an isometry. Because S is normal, there is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block is a 1-by-1 matrix or a 2-by-2 matrix of the form

9.37
$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b > 0$ (by 9.34).

If λ is an entry in a 1-by-1 matrix along the diagonal of the matrix of S (with respect to the basis mentioned above), then there is a basis vector e_j such that $Se_j = \lambda e_j$. Because S is an isometry, this implies that $|\lambda| = 1$. Thus $\lambda = 1$ or $\lambda = -1$, because these are the only real numbers with absolute value 1.

Now consider a 2-by-2 matrix of the form 9.37 along the diagonal of the matrix of S. There are basis vectors e_j, e_{j+1} such that

$$Se_j = ae_j + be_{j+1}.$$

Thus

$$1 = \|e_j\|^2 = \|Se_j\|^2 = a^2 + b^2.$$

The equation above, along with the condition $b > 0$, implies that there exists a number $\theta \in (0, \pi)$ such that $a = \cos\theta$ and $b = \sin\theta$. Thus the matrix 9.37 has the required form, completing the proof in this direction.

Conversely, now suppose (b) holds, so there is an orthonormal basis of V with respect to which the matrix of S has the form required by the theorem. Thus there is a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a subspace of V of dimension 1 or 2. Furthermore, any two vectors belonging to distinct U's are orthogonal, and each $S|_{U_j}$ is an isometry mapping U_j into U_j. If $v \in V$, we can write

$$v = u_1 + \cdots + u_m,$$

where each u_j is in U_j. Applying S to the equation above and then taking norms gives

$$\begin{aligned}
\|Sv\|^2 &= \|Su_1 + \cdots + Su_m\|^2 \\
&= \|Su_1\|^2 + \cdots + \|Su_m\|^2 \\
&= \|u_1\|^2 + \cdots + \|u_m\|^2 \\
&= \|v\|^2.
\end{aligned}$$

Thus S is an isometry, and hence (a) holds. ■

EXERCISES 9.B

1 Suppose $S \in \mathcal{L}(\mathbf{R}^3)$ is an isometry. Prove that there exists a nonzero vector $x \in \mathbf{R}^3$ such that $S^2 x = x$.

2 Prove that every isometry on an odd-dimensional real inner product space has 1 or -1 as an eigenvalue.

3 Suppose V is a real inner product space. Show that

$$\langle u + iv, x + iy \rangle = \langle u, x \rangle + \langle v, y \rangle + \big(\langle v, x \rangle - \langle u, y \rangle\big)i$$

for $u, v, x, y \in V$ defines a complex inner product on $V_{\mathbf{C}}$.

4 Suppose V is a real inner product space and $T \in \mathcal{L}(V)$ is self-adjoint. Show that $T_{\mathbf{C}}$ is a self-adjoint operator on the inner product space $V_{\mathbf{C}}$ defined by the previous exercise.

5 Use the previous exercise to give a proof of the Real Spectral Theorem (7.29) via complexification and the Complex Spectral Theorem (7.24).

6 Give an example of an operator T on an inner product space such that T has an invariant subspace whose orthogonal complement is not invariant under T.
[*The exercise above shows that 9.30 can fail without the hypothesis that T is normal.*]

7 Suppose $T \in \mathcal{L}(V)$ and T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix}$$

with respect to some basis of V. For $j = 1, \dots, m$, let T_j be the operator on V whose matrix with respect to the same basis is a block diagonal matrix with blocks the same size as in the matrix above, with A_j in the j^{th} block, and with all the other blocks on the diagonal equal to identity matrices (of the appropriate size). Prove that $T = T_1 \cdots T_m$.

8 Suppose D is the differentiation operator on the vector space V in Exercise 21 in Section 7.A. Find an orthonormal basis of V such that the matrix of the normal operator D has the form promised by 9.34.

CHAPTER

10

British mathematician and pioneer computer scientist Ada Lovelace (1815–1852), as painted by Alfred Chalon in this 1840 portrait.

Trace and Determinant

Throughout this book our emphasis has been on linear maps and operators rather than on matrices. In this chapter we pay more attention to matrices as we define the trace and determinant of an operator and then connect these notions to the corresponding notions for matrices. The book concludes with an explanation of the important role played by determinants in the theory of volume and integration.

Our assumptions for this chapter are as follows:

10.1 Notation $\mathbf{F}$, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- change of basis and its effect upon the matrix of an operator
- trace of an operator and of a matrix
- determinant of an operator and of a matrix
- determinants and volume

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6_10

10.A Trace

For our study of the trace and determinant, we will need to know how the matrix of an operator changes with a change of basis. Thus we begin this chapter by developing the necessary material about change of basis.

Change of Basis

With respect to every basis of V, the matrix of the identity operator $I \in \mathcal{L}(V)$ is the diagonal matrix with 1's on the diagonal and 0's elsewhere. We also use the symbol I for the name of this matrix, as shown in the next definition.

10.2 Definition ***identity matrix, I***

Suppose n is a positive integer. The n-by-n diagonal matrix

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

is called the ***identity matrix*** and is denoted I.

Note that we use the symbol I to denote the identity operator (on all vector spaces) and the identity matrix (of all possible sizes). You should always be able to tell from the context which particular meaning of I is intended. For example, consider the equation $\mathcal{M}(I) = I$; on the left side I denotes the identity operator, and on the right side I denotes the identity matrix.

If A is a square matrix (with entries in $\mathbf{F}$, as usual) with the same size as I, then $AI = IA = A$, as you should verify.

10.3 Definition ***invertible, inverse,*** A^{-1}

A square matrix A is called ***invertible*** if there is a square matrix B of the same size such that $AB = BA = I$; we call B the ***inverse*** of A and denote it by A^{-1}.

*Some mathematicians use the terms **nonsingular**, which means the same as invertible, and **singular**, which means the same as noninvertible.*

The same proof as used in 3.54 shows that if A is an invertible square matrix, then there is a unique matrix B such that $AB = BA = I$ (and thus the notation $B = A^{-1}$ is justified).

In Section 3.C we defined the matrix of a linear map from one vector space to another with respect to two bases—one basis of the first vector space and another basis of the second vector space. When we study operators, which are linear maps from a vector space to itself, we almost always use the same basis for both vector spaces (after all, the two vector spaces in question are equal). Thus we usually refer to the matrix of an operator with respect to a basis and display at most one basis because we are using one basis in two capacities.

The next result is one of the unusual cases in which we use two different bases even though we have operators from a vector space to itself. It is just a convenient restatement of 3.43 (with U and W both equal to V), but now we are being more careful to include the various bases explicitly in the notation. The result below holds because we defined matrix multiplication to make it true—see 3.43 and the material preceding it.

10.4 The matrix of the product of linear maps

Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ and $w_1, \dots, w_n$ are all bases of V. Suppose $S, T \in \mathcal{L}(V)$. Then

$$\mathcal{M}\big(ST, (u_1, \dots, u_n), (w_1, \dots, w_n)\big) =$$
$$\mathcal{M}\big(S, (v_1, \dots, v_n), (w_1, \dots, w_n)\big)\mathcal{M}\big(T, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

The next result deals with the matrix of the identity operator I with respect to two different bases. Note that the k^{th} column of the matrix $\mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$ consists of the scalars needed to write u_k as a linear combination of $v_1, \dots, v_n$.

10.5 Matrix of the identity with respect to two bases

Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then the matrices $\mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$ and $\mathcal{M}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big)$ are invertible, and each is the inverse of the other.

Proof In 10.4, replace w_j with u_j, and replace S and T with I, getting

$$I = \mathcal{M}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big)\mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

Now interchange the roles of the u's and v's, getting

$$I = \mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)\mathcal{M}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big).$$

These two equations give the desired result. ■

10.6 Example Consider the bases $(4,2),(5,3)$ and $(1,0),(0,1)$ of $\mathbf{F}^2$. Obviously

$$\mathcal{M}\Big(I,\big((4,2),(5,3)\big),\big((1,0),(0,1)\big)\Big)=\begin{pmatrix}4&5\\2&3\end{pmatrix},$$

because $I(4,2)=4(1,0)+2(0,1)$ and $I(5,3)=5(1,0)+3(0,1)$.

The inverse of the matrix above is

$$\begin{pmatrix}\frac{3}{2}&-\frac{5}{2}\\-1&2\end{pmatrix},$$

as you should verify. Thus 10.5 implies that

$$\mathcal{M}\Big(I,\big((1,0),(0,1)\big),\big((4,2),(5,3)\big)\Big)=\begin{pmatrix}\frac{3}{2}&-\frac{5}{2}\\-1&2\end{pmatrix}.$$

Now we can see how the matrix of T changes when we change bases. In the result below, we have two different bases of V. Recall that the notation $\mathcal{M}\big(T,(u_1,\dots,u_n)\big)$ is shorthand for $\mathcal{M}\big(T,(u_1,\dots,u_n),(u_1,\dots,u_n)\big)$

10.7 Change of basis formula

Suppose $T\in\mathcal{L}(V)$. Let $u_1,\dots,u_n$ and $v_1,\dots,v_n$ be bases of V. Let $A=\mathcal{M}\big(I,(u_1,\dots,u_n),(v_1,\dots,v_n)\big)$. Then

$$\mathcal{M}\big(T,(u_1,\dots,u_n)\big)=A^{-1}\mathcal{M}\big(T,(v_1,\dots,v_n)\big)A.$$

Proof In 10.4, replace w_j with u_j and replace S with I, getting

10.8 $$\mathcal{M}\big(T,(u_1,\dots,u_n)\big)=A^{-1}\mathcal{M}\big(T,(u_1,\dots,u_n),(v_1,\dots,v_n)\big),$$

where we have used 10.5.

Again use 10.4, this time replacing w_j with v_j. Also replace T with I and replace S with T, getting

$$\mathcal{M}\big(T,(u_1,\dots,u_n),(v_1,\dots,v_n)\big)=\mathcal{M}\big(T,(v_1,\dots,v_n)\big)A.$$

Substituting the equation above into 10.8 gives the desired result. ■

Trace: A Connection Between Operators and Matrices

Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Let $n = \dim V$. Recall that we defined the multiplicity of λ to be the dimension of the generalized eigenspace $G(\lambda, T)$ (see 8.24) and that this multiplicity equals $\dim \operatorname{null}(T - \lambda I)^n$ (see 8.11). Recall also that if V is a complex vector space, then the sum of the multiplicities of all the eigenvalues of T equals n (see 8.26).

In the definition below, the sum of the eigenvalues "with each eigenvalue repeated according to its multiplicity" means that if $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with multiplicities $d_1, \dots, d_m$, then the sum is

$$d_1\lambda_1 + \cdots + d_m\lambda_m.$$

Or if you prefer to list the eigenvalues with each repeated according to its multiplicity, then the eigenvalues could be denoted $\lambda_1, \dots, \lambda_n$ (where the index n equals $\dim V$) and the sum is

$$\lambda_1 + \cdots + \lambda_n.$$

10.9 Definition ***trace of an operator***

Suppose $T \in \mathcal{L}(V)$.

- If $\mathbf{F} = \mathbf{C}$, then the ***trace*** of T is the sum of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the ***trace*** of T is the sum of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The trace of T is denoted by $\operatorname{trace} T$.

10.10 Example Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator whose matrix is

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

Then the eigenvalues of T are 1, $2 + 3i$, and $2 - 3i$, each with multiplicity 1, as you can verify. Computing the sum of the eigenvalues, we find that $\operatorname{trace} T = 1 + (2 + 3i) + (2 - 3i)$; in other words, $\operatorname{trace} T = 5$.

The trace has a close connection with the characteristic polynomial. Suppose $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with each eigenvalue repeated according to its multiplicity. Then by definition (see 8.34 and 9.21), the characteristic polynomial of T equals

$$(z - \lambda_1) \cdots (z - \lambda_n).$$

Expanding the polynomial above, we can write the characteristic polynomial of T in the form

10.11 $$z^n - (\lambda_1 + \cdots + \lambda_n) z^{n-1} + \cdots + (-1)^n (\lambda_1 \cdots \lambda_n).$$

The expression above immediately leads to the following result.

10.12 Trace and characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then trace T equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T.

Most of the rest of this section is devoted to discovering how to compute trace T from the matrix of T (with respect to an arbitrary basis).

Let's start with the easiest situation. Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and we choose a basis of V as in 8.29. With respect to that basis, T has an upper-triangular matrix with the diagonal of the matrix containing precisely the eigenvalues of T, each repeated according to its multiplicity. Thus trace T equals the sum of the diagonal entries of $\mathcal{M}(T)$ with respect to that basis.

The same formula works for the operator $T \in \mathcal{L}(\mathbf{C}^3)$ in Example 10.10 whose trace equals 5. In that example, the matrix is not in upper-triangular form. However, the sum of the diagonal entries of the matrix in that example equals 5, which is the trace of the operator T.

At this point you should suspect that trace T equals the sum of the diagonal entries of the matrix of T with respect to an arbitrary basis. Remarkably, this suspicion turns out to be true. To prove it, we start by making the following definition.

10.13 **Definition** ***trace of a matrix***

The ***trace*** of a square matrix A, denoted trace A, is defined to be the sum of the diagonal entries of A.

Now we have defined the trace of an operator and the trace of a square matrix, using the same word "trace" in two different contexts. This would be bad terminology unless the two concepts turn out to be essentially the same. As we will see, it is indeed true that $\operatorname{trace} T = \operatorname{trace} \mathcal{M}\big(T, (v_1, \dots, v_n)\big)$, where $v_1, \dots, v_n$ is an arbitrary basis of V. We will need the following result for the proof.

10.14 Trace of AB equals trace of BA

If A and B are square matrices of the same size, then

$$\operatorname{trace}(AB) = \operatorname{trace}(BA).$$

Proof Suppose

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}, \quad B = \begin{pmatrix} B_{1,1} & \dots & B_{1,n} \\ \vdots & & \vdots \\ B_{n,1} & \dots & B_{n,n} \end{pmatrix}.$$

The j^{th} term on the diagonal of AB equals

$$\sum_{k=1}^{n} A_{j,k} B_{k,j}.$$

Thus

$$\begin{aligned} \operatorname{trace}(AB) &= \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k} B_{k,j} \\ &= \sum_{k=1}^{n} \sum_{j=1}^{n} B_{k,j} A_{j,k} \\ &= \sum_{k=1}^{n} k^{\text{th}} \text{ term on the diagonal of } BA \\ &= \operatorname{trace}(BA), \end{aligned}$$

as desired. ■

Now we can prove that the sum of the diagonal entries of the matrix of an operator is independent of the basis with respect to which the matrix is computed.

10.15 Trace of matrix of operator does not depend on basis

Let $T \in \mathcal{L}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then

$$\operatorname{trace} \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) = \operatorname{trace} \mathcal{M}\big(T, (v_1, \ldots, v_n)\big).$$

Proof Let $A = \mathcal{M}\big(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)\big)$. Then

$$\begin{aligned} \operatorname{trace} \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) &= \operatorname{trace}\Big(A^{-1}\big(\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)A\big)\Big) \\ &= \operatorname{trace}\Big(\big(\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)A\big)A^{-1}\Big) \\ &= \operatorname{trace} \mathcal{M}\big(T, (v_1, \ldots, v_n)\big), \end{aligned}$$

where the first equality comes from 10.7 and the second equality follows from 10.14. The third equality completes the proof. ■

The result below, which is the most important result in this section, states that the trace of an operator equals the sum of the diagonal entries of the matrix of the operator. This theorem does not specify a basis because, by the result above, the sum of the diagonal entries of the matrix of an operator is the same for every choice of basis.

10.16 Trace of an operator equals trace of its matrix

Suppose $T \in \mathcal{L}(V)$. Then $\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T)$.

Proof As noted above, $\operatorname{trace} \mathcal{M}(T)$ is independent of which basis of V we choose (by 10.15). Thus to show that

$$\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T)$$

for every basis of V, we need only show that the equation above holds for some basis of V.

As we have already discussed, if V is a complex vector space, then choosing the basis as in 8.29 gives the desired result. If V is a real vector space, then applying the complex case to the complexification $T_{\mathbf{C}}$ (which is used to define $\operatorname{trace} T$) gives the desired result. ■

If we know the matrix of an operator on a complex vector space, the result above allows us to find the sum of all the eigenvalues without finding any of the eigenvalues, as shown by the next example.

10.17 Example Consider the operator on $\mathbf{C}^5$ whose matrix is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

No one can find an exact formula for any of the eigenvalues of this operator. However, we do know that the sum of the eigenvalues equals 0, because the sum of the diagonal entries of the matrix above equals 0.

We can use 10.16 to give easy proofs of some useful properties about traces of operators by shifting to the language of traces of matrices, where certain properties have already been proved or are obvious. The proof of the next result is an example of this technique. The eigenvalues of $S + T$ are not, in general, formed from adding together eigenvalues of S and eigenvalues of T. Thus the next result would be difficult to prove without using 10.16.

10.18 Trace is additive

Suppose $S, T \in \mathcal{L}(V)$. Then $\operatorname{trace}(S + T) = \operatorname{trace} S + \operatorname{trace} T$.

Proof Choose a basis of V. Then

$$\begin{aligned} \operatorname{trace}(S + T) &= \operatorname{trace} \mathcal{M}(S + T) \\ &= \operatorname{trace}\bigl(\mathcal{M}(S) + \mathcal{M}(T)\bigr) \\ &= \operatorname{trace} \mathcal{M}(S) + \operatorname{trace} \mathcal{M}(T) \\ &= \operatorname{trace} S + \operatorname{trace} T, \end{aligned}$$

where again the first and last equalities come from 10.16; the third equality is obvious from the definition of the trace of a matrix. ■

The techniques we have developed have the following curious consequence. A generalization of this result to infinite-dimensional vector spaces has important consequences in modern physics, particularly in quantum theory.

The statement of the next result does not involve traces, although the short proof uses traces. Whenever something like this happens in mathematics, we can be sure that a good definition lurks in the background.

10.19 The identity is not the difference of ST and TS

There do not exist operators $S, T \in \mathcal{L}(V)$ such that $ST - TS = I$.

Proof Suppose $S, T \in \mathcal{L}(V)$. Choose a basis of V. Then

$$\begin{aligned}\operatorname{trace}(ST - TS) &= \operatorname{trace}(ST) - \operatorname{trace}(TS)\\ &= \operatorname{trace}\mathcal{M}(ST) - \operatorname{trace}\mathcal{M}(TS)\\ &= \operatorname{trace}\bigl(\mathcal{M}(S)\mathcal{M}(T)\bigr) - \operatorname{trace}\bigl(\mathcal{M}(T)\mathcal{M}(S)\bigr)\\ &= 0,\end{aligned}$$

where the first equality comes from 10.18, the second equality comes from 10.16, the third equality comes from 3.43, and the fourth equality comes from 10.14. Clearly the trace of I equals $\dim V$, which is not 0. Because $ST - TS$ and I have different traces, they cannot be equal. ■

EXERCISES 10.A

1 Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V. Prove that the matrix $\mathcal{M}\bigl(T, (v_1, \dots, v_n)\bigr)$ is invertible if and only if T is invertible.

2 Suppose A and B are square matrices of the same size and $AB = I$. Prove that $BA = I$.

3 Suppose $T \in \mathcal{L}(V)$ has the same matrix with respect to every basis of V. Prove that T is a scalar multiple of the identity operator.

4 Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Let $T \in \mathcal{L}(V)$ be the operator such that $Tv_k = u_k$ for $k = 1, \dots, n$. Prove that

$$\mathcal{M}\bigl(T, (v_1, \dots, v_n)\bigr) = \mathcal{M}\bigl(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\bigr).$$

5 Suppose B is a square matrix with complex entries. Prove that there exists an invertible square matrix A with complex entries such that $A^{-1}BA$ is an upper-triangular matrix.

6 Give an example of a real vector space V and $T \in \mathcal{L}(V)$ such that $\operatorname{trace}(T^2) < 0$.

7 Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and V has a basis consisting of eigenvectors of T. Prove that $\operatorname{trace}(T^2) \geq 0$.

8 Suppose V is an inner product space and $v, w \in V$. Define $T \in \mathcal{L}(V)$ by $Tu = \langle u, v\rangle w$. Find a formula for trace T.

9 Suppose $P \in \mathcal{L}(V)$ satisfies $P^2 = P$. Prove that

$$\operatorname{trace} P = \dim \operatorname{range} P.$$

10 Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Prove that

$$\operatorname{trace} T^* = \overline{\operatorname{trace} T}.$$

11 Suppose V is an inner product space. Suppose $T \in \mathcal{L}(V)$ is a positive operator and trace $T = 0$. Prove that $T = 0$.

12 Suppose V is an inner product space and $P, Q \in \mathcal{L}(V)$ are orthogonal projections. Prove that $\operatorname{trace}(PQ) \geq 0$.

13 Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator whose matrix is

$$\begin{pmatrix} 51 & -12 & -21 \\ 60 & -40 & -28 \\ 57 & -68 & 1 \end{pmatrix}.$$

Someone tells you (accurately) that -48 and 24 are eigenvalues of T. Without using a computer or writing anything down, find the third eigenvalue of T.

14 Suppose $T \in \mathcal{L}(V)$ and $c \in \mathbf{F}$. Prove that $\operatorname{trace}(cT) = c \operatorname{trace} T$.

15 Suppose $S, T \in \mathcal{L}(V)$. Prove that $\operatorname{trace}(ST) = \operatorname{trace}(TS)$.

16 Prove or give a counterexample: if $S, T \in \mathcal{L}(V)$, then $\operatorname{trace}(ST) = (\operatorname{trace} S)(\operatorname{trace} T)$.

17 Suppose $T \in \mathcal{L}(V)$ is such that $\operatorname{trace}(ST) = 0$ for all $S \in \mathcal{L}(V)$. Prove that $T = 0$.

18 Suppose V is an inner product space with orthonormal basis $e_1, \dots, e_n$ and $T \in \mathcal{L}(V)$. Prove that

$$\operatorname{trace}(T^*T) = \|Te_1\|^2 + \cdots + \|Te_n\|^2.$$

Conclude that the right side of the equation above is independent of which orthonormal basis $e_1, \dots, e_n$ is chosen for V.

19 Suppose V is an inner product space. Prove that

$$\langle S, T \rangle = \operatorname{trace}(ST^*)$$

defines an inner product on $\mathcal{L}(V)$.

20 Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, repeated according to multiplicity. Suppose

$$\begin{pmatrix} A_{1,1} & \ldots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \ldots & A_{n,n} \end{pmatrix}$$

is the matrix of T with respect to some orthonormal basis of V. Prove that

$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \leq \sum_{k=1}^{n} \sum_{j=1}^{n} |A_{j,k}|^2.$$

21 Suppose V is an inner product space. Suppose $T \in \mathcal{L}(V)$ and

$$\|T^* v\| \leq \|T v\|$$

for every $v \in V$. Prove that T is normal.
[*The exercise above fails on infinite-dimensional inner product spaces, leading to what are called hyponormal operators, which have a well-developed theory.*]

10.B *Determinant*

Determinant of an Operator

Now we are ready to define the determinant of an operator. Notice that the definition below mimics the approach we took when defining the trace, with the product of the eigenvalues replacing the sum of the eigenvalues.

10.20 Definition ***determinant of an operator,*** $\det T$

Suppose $T \in \mathcal{L}(V)$.

- If $\mathbf{F} = \mathbf{C}$, then the ***determinant*** of T is the product of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the ***determinant*** of T is the product of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The determinant of T is denoted by $\det T$.

If $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with multiplicities $d_1, \dots, d_m$, then the definition above implies

$$\det T = \lambda_1^{d_1} \cdots \lambda_m^{d_m}.$$

Or if you prefer to list the eigenvalues with each repeated according to its multiplicity, then the eigenvalues could be denoted $\lambda_1, \dots, \lambda_n$ (where the index n equals $\dim V$) and the definition above implies

$$\det T = \lambda_1 \cdots \lambda_n.$$

10.21 Example Suppose $T \in \mathcal{L}(\mathbf{C}^3)$ is the operator whose matrix is

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

Then the eigenvalues of T are 1, $2 + 3i$, and $2 - 3i$, each with multiplicity 1, as you can verify. Computing the product of the eigenvalues, we find that $\det T = 1 \cdot (2 + 3i) \cdot (2 - 3i)$; in other words, $\det T = 13$.

The determinant has a close connection with the characteristic polynomial. Suppose $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with each eigenvalue repeated according to its multiplicity. Then the expression for the characteristic polynomial of T given by 10.11 gives the following result.

10.22 Determinant and characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\det T$ equals $(-1)^n$ times the constant term of the characteristic polynomial of T.

Combining the result above and 10.12, we have the following result.

10.23 Characteristic polynomial, trace, and determinant

Suppose $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T can be written as

$$z^n - (\operatorname{trace} T)z^{n-1} + \cdots + (-1)^n(\det T).$$

We turn now to some simple but important properties of determinants. Later we will discover how to calculate $\det T$ from the matrix of T (with respect to an arbitrary basis).

The crucial result below has an easy proof due to our definition.

10.24 Invertible is equivalent to nonzero determinant

An operator on V is invertible if and only if its determinant is nonzero.

Proof First suppose V is a complex vector space and $T \in \mathcal{L}(V)$. The operator T is invertible if and only if 0 is not an eigenvalue of T. Clearly this happens if and only if the product of the eigenvalues of T is not 0. Thus T is invertible if and only if $\det T \neq 0$, as desired.

Now consider the case where V is a real vector space and $T \in \mathcal{L}(V)$. Again, T is invertible if and only if 0 is not an eigenvalue of T, which happens if and only if 0 is not an eigenvalue of $T_{\mathbf{C}}$ (because $T_{\mathbf{C}}$ and T have the same real eigenvalues by 9.11). Thus again we see that T is invertible if and only if $\det T \neq 0$. ■

Some textbooks take the result below as the definition of the characteristic polynomial and then have our definition of the characteristic polynomial as a consequence.

10.25 Characteristic polynomial of T equals $\det(zI - T)$

Suppose $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T equals $\det(zI - T)$.

Proof First suppose V is a complex vector space. If $\lambda, z \in \mathbf{C}$, then λ is an eigenvalue of T if and only if $z - \lambda$ is an eigenvalue of $zI - T$, as can be seen from the equation

$$-(T - \lambda I) = (zI - T) - (z - \lambda)I.$$

Raising both sides of this equation to the $\dim V$ power and then taking null spaces of both sides shows that the multiplicity of λ as an eigenvalue of T equals the multiplicity of $z - \lambda$ as an eigenvalue of $zI - T$.

Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of T, repeated according to multiplicity. Thus for $z \in \mathbf{C}$, the paragraph above shows that the eigenvalues of $zI - T$ are $z - \lambda_1, \dots, z - \lambda_n$, repeated according to multiplicity. The determinant of $zI - T$ is the product of these eigenvalues. In other words,

$$\det(zI - T) = (z - \lambda_1) \cdots (z - \lambda_n).$$

The right side of the equation above is, by definition, the characteristic polynomial of T, completing the proof when V is a complex vector space.

Now suppose V is a real vector space. Applying the complex case to $T_{\mathbf{C}}$ gives the desired result. ■

Determinant of a Matrix

Our next task is to discover how to compute $\det T$ from the matrix of T (with respect to an arbitrary basis). Let's start with the easiest situation. Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and we choose a basis of V as in 8.29. With respect to that basis, T has an upper-triangular matrix with the diagonal of the matrix containing precisely the eigenvalues of T, each repeated according to its multiplicity. Thus $\det T$ equals the product of the diagonal entries of $\mathcal{M}(T)$ with respect to that basis.

When dealing with the trace in the previous section, we discovered that the formula (trace = sum of diagonal entries) that worked for the upper-triangular matrix given by 8.29 also worked with respect to an arbitrary basis. Could that also work for determinants? In other words, is the determinant of an operator equal to the product of the diagonal entries of the matrix of the operator with respect to an arbitrary basis?

Unfortunately, the determinant is more complicated than the trace. In particular, $\det T$ need not equal the product of the diagonal entries of $\mathcal{M}(T)$ with respect to an arbitrary basis. For example, the operator in Example 10.21 has determinant 13 but the product of the diagonal entries of its matrix equals 0.

For each square matrix A, we want to define the determinant of A, denoted $\det A$, so that $\det T = \det \mathcal{M}(T)$ regardless of which basis is used to compute $\mathcal{M}(T)$. We begin our search for the correct definition of the determinant of a matrix by calculating the determinants of some special operators.

10.26 Example Suppose $a_1, \dots, a_n \in \mathbf{F}$. Let

$$A = \begin{pmatrix} 0 & & & & a_n \\ a_1 & 0 & & & \\ & a_2 & 0 & & \\ & & \ddots & \ddots & \\ & & & a_{n-1} & 0 \end{pmatrix};$$

here all entries of the matrix are 0 except for the upper-right corner and along the line just below the diagonal. Suppose $v_1, \dots, v_n$ is a basis of V and $T \in \mathcal{L}(V)$ is such that $\mathcal{M}\big(T, (v_1, \dots, v_n)\big) = A$. Find the determinant of T.

Solution First assume $a_j \neq 0$ for each $j = 1, \dots, n-1$. Note that the list $v_1, Tv_1, T^2v_1, \dots, T^{n-1}v_1$ equals $v_1, a_1v_2, a_1a_2v_3, \dots, a_1 \cdots a_{n-1}v_n$. Thus $v_1, Tv_1, \dots, T^{n-1}v_1$ is linearly independent (because the a's are all nonzero). Hence if p is a monic polynomial with degree at most $n-1$, then $p(T)v_1 \neq 0$. Thus the minimal polynomial of T cannot have degree less than n.

Computing the minimal polynomial is often an efficient method of finding the characteristic polynomial, as is done in this example.

As you should verify, $T^n v_j = a_1 \cdots a_n v_j$ for each j. Thus we have $T^n = a_1 \cdots a_n I$. Hence $z^n - a_1 \cdots a_n$ is the minimal polynomial of T. Because $n = \dim V$ and the characteristic polynomial is a polynomial multiple of the minimal polynomial (9.26), this implies that $z^n - a_1 \cdots a_n$ is also the characteristic polynomial of T.

Thus 10.22 implies that

$$\det T = (-1)^{n-1} a_1 \cdots a_n.$$

If some a_j equals 0, then $Tv_j = 0$ for some j, which implies that 0 is an eigenvalue of T and hence $\det T = 0$. In other words, the formula above also holds if some a_j equals 0.

Thus in order to have $\det T = \det \mathcal{M}(T)$, we will have to make the determinant of the matrix in Example 10.26 equal to $(-1)^{n-1} a_1 \cdots a_n$. However, we do not yet have enough evidence to make a reasonable guess about the proper definition of the determinant of an arbitrary square matrix.

To compute the determinants of a more complicated class of operators, we introduce the notion of permutation.

10.27 Definition ***permutation,*** $\operatorname{perm} n$

- A ***permutation*** of $(1, \ldots, n)$ is a list $(m_1, \ldots, m_n)$ that contains each of the numbers $1, \ldots, n$ exactly once.
- The set of all permutations of $(1, \ldots, n)$ is denoted $\operatorname{perm} n$.

For example, $(2, 3, 4, 5, 1) \in \operatorname{perm} 5$. You should think of an element of $\operatorname{perm} n$ as a rearrangement of the first n integers.

10.28 Example Suppose $a_1, \ldots, a_n \in \mathbf{F}$ and $v_1, \ldots, v_n$ is a basis of V. Consider a permutation $(p_1, \ldots, p_n) \in \operatorname{perm} n$ that can be obtained as follows: break $(1, \ldots, n)$ into lists of consecutive integers and in each list move the first term to the end of that list. For example, taking $n = 9$, the permutation

$$(2, 3, 1, 5, 6, 7, 4, 9, 8)$$

is obtained from $(1, 2, 3)$, $(4, 5, 6, 7)$, $(8, 9)$ by moving the first term of each of these lists to the end, producing $(2, 3, 1)$, $(5, 6, 7, 4)$, $(9, 8)$, and then putting these together to form the permutation displayed above.

Let $T \in \mathcal{L}(V)$ be the operator such that

$$Tv_k = a_k v_{p_k}$$

for $k = 1, \ldots, n$. Find $\det T$.

Solution This generalizes Example 10.26, because if $(p_1, \ldots, p_n)$ is the permutation $(2, 3, \ldots, n, 1)$, then our operator T is the same as the operator T in Example 10.26.

With respect to the basis $v_1, \ldots, v_n$, the matrix of the operator T is a block diagonal matrix

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_M \end{pmatrix},$$

where each block is a square matrix of the form of the matrix in 10.26.

Correspondingly, we can write $V = V_1 \oplus \cdots \oplus V_M$, where each V_j is invariant under T and each $T|_{V_j}$ is of the form of the operator in 10.26. Because $\det T = (\det T|_{V_1}) \cdots (\det T|_{V_M})$ (because the dimensions of the generalized eigenspaces in the V_j add up to $\dim V$), we have

$$\det T = (-1)^{n_1 - 1} \cdots (-1)^{n_M - 1} a_1 \cdots a_n,$$

where V_j has dimension n_j (and correspondingly each A_j has size n_j-by-n_j) and we have used the result from 10.26.

The number $(-1)^{n_1 - 1} \cdots (-1)^{n_M - 1}$ that appears above is called the sign of the corresponding permutation $(p_1, \dots, p_n)$, denoted $\operatorname{sign}(p_1, \dots, p_n)$ [this is a temporary definition that we will change to an equivalent definition later, when we define the sign of an arbitrary permutation].

To put this into a form that does not depend on the particular permutation $(p_1, \dots, p_n)$, let $A_{j,k}$ denote the entry in row j, column k, of the matrix A from Example 10.28. Thus

$$A_{j,k} = \begin{cases} 0 & \text{if } j \neq p_k; \\ a_k & \text{if } j = p_k. \end{cases}$$

Example 10.28 shows that we want

10.29 $$\det A = \sum_{(m_1, \dots, m_n) \in \operatorname{perm} n} \bigl(\operatorname{sign}(m_1, \dots, m_n)\bigr) A_{m_1,1} \cdots A_{m_n,n};$$

note that each summand is 0 except the one corresponding to the permutation $(p_1, \dots, p_n)$ [which is why it does not matter that the sign of the other permutations is not yet defined].

We can now guess that $\det A$ should be defined by 10.29 for an arbitrary square matrix A. This will turn out to be correct. We will now dispense with the motivation and begin the more formal approach. First we will need to define the sign of an arbitrary permutation.

10.30 Definition ***sign of a permutation***

- The ***sign*** of a permutation $(m_1, \dots, m_n)$ is defined to be 1 if the number of pairs of integers (j, k) with $1 \leq j < k \leq n$ such that j appears after k in the list $(m_1, \dots, m_n)$ is even and -1 if the number of such pairs is odd.

- In other words, the sign of a permutation equals 1 if the natural order has been changed an even number of times and equals -1 if the natural order has been changed an odd number of times.

10.31 Example ***sign of permutation***

- The only pair of integers (j, k) with $j < k$ such that j appears after k in the list $(2, 1, 3, 4)$ is $(1, 2)$. Thus the permutation $(2, 1, 3, 4)$ has sign -1.
- In the permutation $(2, 3, \dots, n, 1)$, the only pairs (j, k) with $j < k$ that appear with changed order are $(1, 2), (1, 3), \dots, (1, n)$; because we have $n - 1$ such pairs, the sign of this permutation equals $(-1)^{n-1}$ (note that the same quantity appeared in Example 10.26).

The next result shows that interchanging two entries of a permutation changes the sign of the permutation.

10.32 Interchanging two entries in a permutation

Interchanging two entries in a permutation multiplies the sign of the permutation by -1.

Proof Suppose we have two permutations, where the second permutation is obtained from the first by interchanging two entries. If the two interchanged entries were in their natural order in the first permutation, then they no longer are in the second permutation, and vice versa, for a net change (so far) of 1 or -1 (both odd numbers) in the number of pairs not in their natural order.

*Some texts use the term **signum**, which means the same as sign.*

Consider each entry between the two interchanged entries. If an intermediate entry was originally in the natural order with respect to both interchanged entries, then it is now in the natural order with respect to neither interchanged entry. Similarly, if an intermediate entry was originally in the natural order with respect to neither of the interchanged entries, then it is now in the natural order with respect to both interchanged entries. If an intermediate entry was originally in the natural order with respect to exactly one of the interchanged entries, then that is still true. Thus the net change for each intermediate entry in the number of pairs not in their natural order is 2, -2, or 0 (all even numbers).

For all the other entries, there is no change in the number of pairs not in their natural order. Thus the total net change in the number of pairs not in their natural order is an odd number. Thus the sign of the second permutation equals -1 times the sign of the first permutation. ■

Our motivation for the next definition comes from 10.29.

10.33 Definition ***determinant of a matrix,*** det A

Suppose A is an n-by-n matrix

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}.$$

The ***determinant*** of A, denoted det A, is defined by

$$\det A = \sum_{(m_1,\dots,m_n)\in \operatorname{perm} n} \bigl(\operatorname{sign}(m_1,\dots,m_n)\bigr) A_{m_1,1} \cdots A_{m_n,n}.$$

10.34 Example ***determinants***

- If A is the 1-by-1 matrix $[A_{1,1}]$, then $\det A = A_{1,1}$, because perm 1 has only one element, namely (1), which has sign 1.
- Clearly perm 2 has only two elements, namely $(1, 2)$, which has sign 1, and $(2, 1)$, which has sign -1. Thus
$$\det \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} = A_{1,1}A_{2,2} - A_{2,1}A_{1,2}.$$

The set perm 3 *contains six elements. In general,* perm n *contains* $n!$ *elements. Note that* $n!$ *rapidly grows large as n increases.*

To make sure you understand this process, you should now find the formula for the determinant of an arbitrary 3-by-3 matrix using just the definition given above.

10.35 Example Compute the determinant of an upper-triangular matrix

$$A = \begin{pmatrix} A_{1,1} & & * \\ & \ddots & \\ 0 & & A_{n,n} \end{pmatrix}.$$

Solution The permutation $(1, 2, \dots, n)$ has sign 1 and thus contributes a term of $A_{1,1} \cdots A_{n,n}$ to the sum defining det A in 10.33. Any other permutation $(m_1, \dots, m_n) \in \operatorname{perm} n$ contains at least one entry m_j with $m_j > j$, which means that $A_{m_j,j} = 0$ (because A is upper triangular). Thus all the other terms in the sum in 10.33 make no contribution.

Hence $\det A = A_{1,1} \cdots A_{n,n}$. In other words, the determinant of an upper-triangular matrix equals the product of the diagonal entries.

Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and we choose a basis of V as in 8.29. With respect to that basis, T has an upper-triangular matrix with the diagonal of the matrix containing precisely the eigenvalues of T, each repeated according to its multiplicity. Thus Example 10.35 tells us that $\det T = \det \mathcal{M}(T)$, where the matrix is with respect to that basis.

Our goal is to prove that $\det T = \det \mathcal{M}(T)$ for every basis of V, not just the basis from 8.29. To do this, we will need to develop some properties of determinants of matrices. The result below is the first of the properties we will need.

10.36 Interchanging two columns in a matrix

Suppose A is a square matrix and B is the matrix obtained from A by interchanging two columns. Then

$$\det A = -\det B.$$

Proof Think of the sum defining $\det A$ in 10.33 and the corresponding sum defining $\det B$. The same products of $A_{j,k}$'s appear in both sums, although they correspond to different permutations. The permutation corresponding to a given product of $A_{j,k}$'s when computing $\det B$ is obtained by interchanging two entries in the corresponding permutation when computing $\det A$, thus multiplying the sign of the permutation by -1 (see 10.32). Hence we see that $\det A = -\det B$. ■

If $T \in \mathcal{L}(V)$ and the matrix of T (with respect to some basis) has two equal columns, then T is not injective and hence $\det T = 0$. Although this comment makes the next result plausible, it cannot be used in the proof, because we do not yet know that $\det T = \det \mathcal{M}(T)$ for every choice of basis.

10.37 Matrices with two equal columns

If A is a square matrix that has two equal columns, then $\det A = 0$.

Proof Suppose A is a square matrix that has two equal columns. Interchanging the two equal columns of A gives the original matrix A. Thus from 10.36 (with $B = A$), we have

$$\det A = -\det A,$$

which implies that $\det A = 0$. ■

Recall from 3.44 that if A is an n-by-n matrix

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix},$$

then we can think of the k^{th} column of A as an n-by-1 matrix denoted $A_{\cdot,k}$:

$$A_{\cdot,k} = \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{n,k} \end{pmatrix}.$$

Some books define the determinant to be the function defined on the square matrices that is linear as a function of each column separately and that satisfies 10.38 and $\det I = 1$. *To prove that such a function exists and that it is unique takes a nontrivial amount of work.*

Note that $A_{j,k}$, with two subscripts, denotes an entry of A, whereas $A_{\cdot,k}$, with a dot as a placeholder and one subscript, denotes a column of A. This notation allows us to write A in the form

$$\begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,n} \end{pmatrix},$$

which will be useful.

The next result shows that a permutation of the columns of a matrix changes the determinant by a factor of the sign of the permutation.

10.38 Permuting the columns of a matrix

Suppose $A = \begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,n} \end{pmatrix}$ is an n-by-n matrix and $(m_1, \dots, m_n)$ is a permutation. Then

$$\det\begin{pmatrix} A_{\cdot,m_1} & \dots & A_{\cdot,m_n} \end{pmatrix} = \bigl(\operatorname{sign}(m_1, \dots, m_n)\bigr) \det A.$$

Proof We can transform the matrix $\begin{pmatrix} A_{\cdot,m_1} & \dots & A_{\cdot,m_n} \end{pmatrix}$ into A through a series of steps. In each step, we interchange two columns and hence multiply the determinant by -1 (see 10.36). The number of steps needed equals the number of steps needed to transform the permutation $(m_1, \dots, m_n)$ into the permutation $(1, \dots, n)$ by interchanging two entries in each step. The proof is completed by noting that the number of such steps is even if $(m_1, \dots, m_n)$ has sign 1, odd if $(m_1, \dots, m_n)$ has sign -1 (this follows from 10.32, along with the observation that the permutation $(1, \dots, n)$ has sign 1). ■

The next result about determinants will also be useful.

10.39 Determinant is a linear function of each column

Suppose k, n are positive integers with $1 \leq k \leq n$. Fix n-by-1 matrices $A_{\cdot,1}, \dots, A_{\cdot,n}$ except $A_{\cdot,k}$. Then the function that takes an n-by-1 column vector $A_{\cdot,k}$ to

$$\det\begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,k} & \dots & A_{\cdot,n} \end{pmatrix}$$

is a linear map from the vector space of n-by-1 matrices with entries in $\mathbf{F}$ to $\mathbf{F}$.

Proof The linearity follows easily from 10.33, where each term in the sum contains precisely one entry from the k^{th} column of A. ■

The result below was first proved in 1812 by French mathematicians Jacques Binet and Augustin-Louis Cauchy.

Now we are ready to prove one of the key properties about determinants of square matrices. This property will enable us to connect the determinant of an operator with the determinant of its matrix. Note that this proof is considerably more complicated than the proof of the corresponding result about the trace (see 10.14).

10.40 Determinant is multiplicative

Suppose A and B are square matrices of the same size. Then

$$\det(AB) = \det(BA) = (\det A)(\det B).$$

Proof Write $A = \begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,n} \end{pmatrix}$, where each $A_{\cdot,k}$ is an n-by-1 column of A. Also write

$$B = \begin{pmatrix} B_{1,1} & \dots & B_{1,n} \\ \vdots & & \vdots \\ B_{n,1} & \dots & B_{n,n} \end{pmatrix} = \begin{pmatrix} B_{\cdot,1} & \dots & B_{\cdot,n} \end{pmatrix},$$

where each $B_{\cdot,k}$ is an n-by-1 column of B. Let e_k denote the n-by-1 matrix that equals 1 in the k^{th} row and 0 elsewhere. Note that $Ae_k = A_{\cdot,k}$ and $Be_k = B_{\cdot,k}$. Furthermore, $B_{\cdot,k} = \sum_{m=1}^{n} B_{m,k} e_m$.

First we will prove $\det(AB) = (\det A)(\det B)$. As we observed earlier (see 3.49), the definition of matrix multiplication easily implies that $AB = \begin{pmatrix} AB_{\cdot,1} & \dots & AB_{\cdot,n} \end{pmatrix}$. Thus

$$\begin{aligned}
\det(AB) &= \det\begin{pmatrix} AB_{\cdot,1} & \dots & AB_{\cdot,n} \end{pmatrix} \\
&= \det\begin{pmatrix} A\left(\sum_{m_1=1}^{n} B_{m_1,1} e_{m_1}\right) & \dots & A\left(\sum_{m_n=1}^{n} B_{m_n,n} e_{m_n}\right) \end{pmatrix} \\
&= \det\begin{pmatrix} \sum_{m_1=1}^{n} B_{m_1,1} A e_{m_1} & \dots & \sum_{m_n=1}^{n} B_{m_n,n} A e_{m_n} \end{pmatrix} \\
&= \sum_{m_1=1}^{n} \cdots \sum_{m_n=1}^{n} B_{m_1,1} \cdots B_{m_n,n} \det\begin{pmatrix} A e_{m_1} & \dots & A e_{m_n} \end{pmatrix},
\end{aligned}$$

where the last equality comes from repeated applications of the linearity of det as a function of one column at a time (10.39). In the last sum above, all terms in which $m_j = m_k$ for some $j \neq k$ can be ignored, because the determinant of a matrix with two equal columns is 0 (by 10.37). Thus instead of summing over all $m_1, \dots, m_n$ with each m_j taking on values $1, \dots, n$, we can sum just over the permutations, where the m_j's have distinct values. In other words,

$$\begin{aligned}
\det(AB) &= \sum_{(m_1,\dots,m_n) \in \operatorname{perm} n} B_{m_1,1} \cdots B_{m_n,n} \det\begin{pmatrix} A e_{m_1} & \dots & A e_{m_n} \end{pmatrix} \\
&= \sum_{(m_1,\dots,m_n) \in \operatorname{perm} n} B_{m_1,1} \cdots B_{m_n,n} \big(\operatorname{sign}(m_1, \dots, m_n)\big) \det A \\
&= (\det A) \sum_{(m_1,\dots,m_n) \in \operatorname{perm} n} \big(\operatorname{sign}(m_1, \dots, m_n)\big) B_{m_1,1} \cdots B_{m_n,n} \\
&= (\det A)(\det B),
\end{aligned}$$

where the second equality comes from 10.38.

In the paragraph above, we proved that $\det(AB) = (\det A)(\det B)$. Interchanging the roles of A and B, we have $\det(BA) = (\det B)(\det A)$. The last equation can be rewritten as $\det(BA) = (\det A)(\det B)$, completing the proof. ■

Note the similarity of the proof of the next result to the proof of the analogous result about the trace (see 10.15).

Now we can prove that the determinant of the matrix of an operator is independent of the basis with respect to which the matrix is computed.

10.41 Determinant of matrix of operator does not depend on basis

Let $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then

$$\det \mathcal{M}\big(T, (u_1, \dots, u_n)\big) = \det \mathcal{M}\big(T, (v_1, \dots, v_n)\big).$$

Proof Let $A = \mathcal{M}\big(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)\big)$. Then

$$\begin{aligned} \det \mathcal{M}\big(T, (u_1, \ldots, u_n)\big) &= \det\Big(A^{-1}\big(\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)A\big)\Big) \\ &= \det\Big(\big(\mathcal{M}\big(T, (v_1, \ldots, v_n)\big)A\big)A^{-1}\Big) \\ &= \det \mathcal{M}\big(T, (v_1, \ldots, v_n)\big), \end{aligned}$$

where the first equality follows from 10.7 and the second equality follows from 10.40. The third equality completes the proof. ■

The result below states that the determinant of an operator equals the determinant of the matrix of the operator. This theorem does not specify a basis because, by the result above, the determinant of the matrix of an operator is the same for every choice of basis.

10.42 Determinant of an operator equals determinant of its matrix

Suppose $T \in \mathcal{L}(V)$. Then $\det T = \det \mathcal{M}(T)$.

Proof As noted above, 10.41 implies that $\det \mathcal{M}(T)$ is independent of which basis of V we choose. Thus to show that $\det T = \det \mathcal{M}(T)$ for every basis of V, we need only show that the result holds for some basis of V.

As we have already discussed, if V is a complex vector space, then choosing a basis of V as in 8.29 gives the desired result. If V is a real vector space, then applying the complex case to the complexification $T_{\mathbf{C}}$ (which is used to define $\det T$) gives the desired result. ■

If we know the matrix of an operator on a complex vector space, the result above allows us to find the product of all the eigenvalues without finding any of the eigenvalues.

10.43 Example Suppose T is the operator on $\mathbf{C}^5$ whose matrix is

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

No one knows an exact formula for any of the eigenvalues of this operator. However, we do know that the product of the eigenvalues equals -3, because the determinant of the matrix above equals -3.

We can use 10.42 to give easy proofs of some useful properties about determinants of operators by shifting to the language of determinants of matrices, where certain properties have already been proved or are obvious. We carry out this procedure in the next result.

10.44 Determinant is multiplicative

Suppose $S, T \in \mathcal{L}(V)$. Then

$$\det(ST) = \det(TS) = (\det S)(\det T).$$

Proof Choose a basis of V. Then

$$\begin{aligned} \det(ST) &= \det \mathcal{M}(ST) \\ &= \det\bigl(\mathcal{M}(S)\mathcal{M}(T)\bigr) \\ &= \bigl(\det \mathcal{M}(S)\bigr)\bigl(\det \mathcal{M}(T)\bigr) \\ &= (\det S)(\det T), \end{aligned}$$

where the first and last equalities come from 10.42 and the third equality comes from 10.40.

In the paragraph above, we proved that $\det(ST) = (\det S)(\det T)$. Interchanging the roles of S and T, we have $\det(TS) = (\det T)(\det S)$. Because multiplication of elements of $\mathbf{F}$ is commutative, the last equation can be rewritten as $\det(TS) = (\det S)(\det T)$, completing the proof. ■

The Sign of the Determinant

We proved the basic results of linear algebra before introducing determinants in this final chapter. Although determinants have value as a research tool in more advanced subjects, they play little role in basic linear algebra (when the subject is done right).

Most applied mathematicians agree that determinants should rarely be used in serious numeric calculations.

Determinants do have one important application in undergraduate mathematics, namely, in computing certain volumes and integrals. In this subsection we interpret the meaning of the sign of the determinant on a real vector space. Then in the final subsection we will use the linear algebra we have learned to make clear the connection between determinants and these applications. Thus we will be dealing with a part of analysis that uses linear algebra.

We will begin with some purely linear algebra results that will also be useful when investigating volumes. Our setting will be inner product spaces. Recall that an isometry on an inner product space is an operator that preserves norms. The next result shows that every isometry has determinant with absolute value 1.

10.45 Isometries have determinant with absolute value 1

Suppose V is an inner product space and $S \in \mathcal{L}(V)$ is an isometry. Then $|\det S| = 1$.

Proof First consider the case where V is a complex inner product space. Then all the eigenvalues of S have absolute value 1 (see the proof of 7.43). Thus the product of the eigenvalues of S, counting multiplicity, has absolute value one. In other words, $|\det S| = 1$, as desired.

Now suppose V is a real inner product space. We present two different proofs in this case.

Proof 1: With respect to the inner product on the complexification $V_{\mathbf{C}}$ given by Exercise 3 in Section 9.B, it is easy to see that $S_{\mathbf{C}}$ is an isometry on $V_{\mathbf{C}}$. Thus by the complex case that we have already done, we have $|\det S_{\mathbf{C}}| = 1$. By definition of the determinant on real vector spaces, we have $\det S = \det S_{\mathbf{C}}$ and thus $|\det S| = 1$, completing the proof.

Proof 2: By 9.36, there is an orthonormal basis of V with respect to which $\mathcal{M}(S)$ is a block diagonal matrix, where each block on the diagonal is a 1-by-1 matrix containing 1 or -1 or a 2-by-2 matrix of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

with $\theta \in (0, \pi)$. Note that the determinant of each 2-by-2 matrix of the form above equals 1 (because $\cos^2\theta + \sin^2\theta = 1$). Thus the determinant of S, which is the product of the determinants of the blocks (see Exercise 6), is the product of 1's and -1's. Hence, $|\det S| = 1$, as desired. ■

The Real Spectral Theorem 7.29 states that a self-adjoint operator T on a real inner product space has an orthonormal basis consisting of eigenvectors. With respect to such a basis, the number of times each eigenvalue appears on the diagonal of $\mathcal{M}(T)$ is its multiplicity. Thus $\det T$ equals the product of its eigenvalues, counting multiplicity (of course, this holds for every operator, self-adjoint or not, on a complex vector space).

Recall that if V is an inner product space and $T \in \mathcal{L}(V)$, then T^*T is a positive operator and hence has a unique positive square root, denoted $\sqrt{T^*T}$ (see 7.35 and 7.36). Because $\sqrt{T^*T}$ is positive, all its eigenvalues are nonnegative (again, see 7.35), and hence $\det \sqrt{T^*T} \geq 0$. These considerations play a role in next example.

10.46 Example Suppose V is a real inner product space and $T \in \mathcal{L}(V)$ is invertible (and thus $\det T$ is either positive or negative). Attach a geometric meaning to the sign of $\det T$.

Solution First we consider an isometry $S \in \mathcal{L}(V)$. By 10.45, the determinant of S equals 1 or -1. Note that

$$\{v \in V : Sv = -v\}$$

We are not formally defining the phrase "reverses direction" because these comments are meant only as an intuitive aid to our understanding.

is the eigenspace $E(-1, S)$. Thinking geometrically, we could say that this is the subspace on which S reverses direction. An examination of proof 2 of 10.45 shows that $\det S = 1$ if this subspace has even dimension and $\det S = -1$ if this subspace has odd dimension.

Returning to our arbitrary invertible operator $T \in \mathcal{L}(V)$, by the Polar Decomposition (7.45) there is an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Now 10.44 tells us that

$$\det T = (\det S)(\det \sqrt{T^*T}).$$

The remarks just before this example pointed out that $\det \sqrt{T^*T} \geq 0$. Thus whether $\det T$ is positive or negative depends on whether $\det S$ is positive or negative. As we saw in the paragraph above, this depends on whether the subspace on which S reverses direction has even or odd dimension.

Because T is the product of S and an operator that never reverses direction (namely, $\sqrt{T^*T}$), we can reasonably say that whether $\det T$ is positive or negative depends on whether T reverses vectors an even or an odd number of times.

Volume

The next result will be a key tool in our investigation of volume. Recall that our remarks before Example 10.46 pointed out that $\det \sqrt{T^*T} \geq 0$.

10.47 $|\det T| = \det \sqrt{T^*T}$

Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Then

$$|\det T| = \det \sqrt{T^*T}.$$

Proof

Another proof of this result is suggested in Exercise 8.

By the Polar Decomposition (7.45), there is an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Thus

$$\begin{aligned} |\det T| &= |\det S| \det \sqrt{T^*T} \\ &= \det \sqrt{T^*T}, \end{aligned}$$

where the first equality follows from 10.44 and the second equality follows from 10.45. ■

Now we turn to the question of volume in $\mathbf{R}^n$. Fix a positive integer n for the rest of this subsection. We will consider only the real inner product space $\mathbf{R}^n$, with its standard inner product.

We would like to assign to each subset Ω of $\mathbf{R}^n$ its n-dimensional volume (when $n = 2$, this is usually called area instead of volume). We begin with boxes, where we have a good intuitive notion of volume.

10.48 Definition ***box***

A ***box*** in $\mathbf{R}^n$ is a set of the form

$$\{(y_1, \dots, y_n) \in \mathbf{R}^n : x_j < y_j < x_j + r_j \text{ for } j = 1, \dots, n\},$$

where $r_1, \dots, r_n$ are positive numbers and $(x_1, \dots, x_n) \in \mathbf{R}^n$. The numbers $r_1, \dots, r_n$ are called the ***side lengths*** of the box.

You should verify that when $n = 2$, a box is a rectangle with sides parallel to the coordinate axes, and that when $n = 3$, a box is a familiar 3-dimensional box with sides parallel to the coordinate axes.

The next definition fits with our intuitive notion of volume, because we define the volume of a box to be the product of the side lengths of the box.

10.49 Definition ***volume of a box***

The ***volume*** of a box B in $\mathbf{R}^n$ with side lengths $r_1, \dots, r_n$ is defined to be $r_1 \cdots r_n$ and is denoted by volume B.

Readers familiar with outer measure will recognize that concept here.

To define the volume of an arbitrary set $\Omega \subset \mathbf{R}^n$, the idea is to write Ω as a subset of a union of many small boxes, then add up the volumes of these small boxes. As we approximate Ω more accurately by unions of small boxes, we get a better estimate of volume Ω.

10.50 Definition ***volume***

Suppose $\Omega \subset \mathbf{R}^n$. Then the ***volume*** of Ω, denoted volume Ω, is defined to be the infimum of

$$\text{volume } B_1 + \text{volume } B_2 + \cdots,$$

where the infimum is taken over all sequences $B_1, B_2, \dots$ of boxes in $\mathbf{R}^n$ whose union contains Ω.

We will work only with an intuitive notion of volume. Our purpose in this book is to understand linear algebra, whereas notions of volume belong to analysis (although volume is intimately connected with determinants, as we will soon see). Thus for the rest of this section we will rely on intuitive notions of volume rather than on a rigorous development, although we shall maintain our usual rigor in the linear algebra parts of what follows. Everything said here about volume will be correct if appropriately interpreted—the intuitive approach used here can be converted into appropriate correct definitions, correct statements, and correct proofs using the machinery of analysis.

10.51 Notation $T(\Omega)$

For T a function defined on a set Ω, define $T(\Omega)$ by

$$T(\Omega) = \{Tx : x \in \Omega\}.$$

For $T \in \mathcal{L}(\mathbf{R}^n)$ and $\Omega \subset \mathbf{R}^n$, we seek a formula for volume $T(\Omega)$ in terms of T and volume Ω. We begin by looking at positive operators.

10.52 Positive operators change volume by factor of determinant

Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ is a positive operator and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume}\, T(\Omega) = (\det T)(\text{volume}\, \Omega).$$

Proof To get a feeling for why this result is true, first consider the special case where $\lambda_1, \dots, \lambda_n$ are positive numbers and $T \in \mathcal{L}(\mathbf{R}^n)$ is defined by

$$T(x_1, \dots, x_n) = (\lambda_1 x_1, \dots, \lambda_n x_n).$$

This operator stretches the j^{th} standard basis vector by a factor of λ_j. If B is a box in $\mathbf{R}^n$ with side lengths $r_1, \dots, r_n$, then $T(B)$ is a box in $\mathbf{R}^n$ with side lengths $\lambda_1 r, \dots, \lambda_n r$. The box $T(B)$ thus has volume $\lambda_1 \cdots \lambda_n r_1 \cdots r_n$, whereas the box Ω has volume $r_1 \cdots r_n$. Note that $\det T = \lambda_1 \cdots \lambda_n$. Thus

$$\text{volume}\, T(B) = (\det T)(\text{volume}\, B)$$

for every box B in $\mathbf{R}^n$. Because the volume of Ω is approximated by sums of volumes of boxes, this implies that volume $T(\Omega) = (\det T)(\text{volume}\, \Omega)$.

Now consider an arbitrary positive operator $T \in \mathcal{L}(\mathbf{R}^n)$. By the Real Spectral Theorem (7.29), there exist an orthonormal basis $e_1, \dots, e_n$ of $\mathbf{R}^n$ and nonnegative numbers $\lambda_1, \dots, \lambda_n$ such that $Te_j = \lambda_j e_j$ for $j = 1, \dots, n$. In the special case where $e_1, \dots, e_n$ is the standard basis of $\mathbf{R}^n$, this operator is the same one as defined in the paragraph above. For an arbitrary orthonormal basis $e_1, \dots, e_n$, this operator has the same behavior as the one in the paragraph above—it stretches the j^{th} basis vector in an orthonormal basis by a factor of λ_j. Your intuition about volume should convince you that volume behaves the same with respect to each orthonormal basis. That intuition, and the special case of the paragraph above, should convince you that T multiplies volume by a factor of $\lambda_1 \cdots \lambda_n$, which again equals $\det T$. ■

Our next tool is the following result, which states that isometries do not change volume.

10.53 An isometry does not change volume

Suppose $S \in \mathcal{L}(\mathbf{R}^n)$ is an isometry and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume}\, S(\Omega) = \text{volume}\, \Omega.$$

Proof For $x, y \in \mathbf{R}^n$, we have

$$\begin{aligned}\|Sx - Sy\| &= \|S(x-y)\| \\ &= \|x-y\|.\end{aligned}$$

In other words, S does not change the distance between points. That property alone may be enough to convince you that S does not change volume.

However, if you need stronger persuasion, consider the complete description of isometries on real inner product spaces provided by 9.36. According to 9.36, S can be decomposed into pieces, each of which is the identity on some subspace (which clearly does not change volume) or multiplication by -1 on some subspace (which again clearly does not change volume) or a rotation on a 2-dimensional subspace (which again does not change volume). Or use 9.36 in conjunction with Exercise 7 in Section 9.B to write S as a product of operators, each of which does not change volume. Either way, you should be convinced that S does not change volume. ■

Now we can prove that an operator $T \in \mathcal{L}(\mathbf{R}^n)$ changes volume by a factor of $|\det T|$. Note the huge importance of the Polar Decomposition in the proof.

10.54 T **changes volume by factor of** $|\det T|$

Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume}\, T(\Omega) = |\det T|(\text{volume}\, \Omega).$$

Proof By the Polar Decomposition (7.45), there is an isometry $S \in \mathcal{L}(V)$ such that

$$T = S\sqrt{T^*T}.$$

If $\Omega \subset \mathbf{R}^n$, then $T(\Omega) = S\big(\sqrt{T^*T}(\Omega)\big)$. Thus

$$\begin{aligned}\text{volume}\, T(\Omega) &= \text{volume}\, S\big(\sqrt{T^*T}(\Omega)\big) \\ &= \text{volume}\, \sqrt{T^*T}(\Omega) \\ &= (\det \sqrt{T^*T})(\text{volume}\, \Omega) \\ &= |\det T|(\text{volume}\, \Omega),\end{aligned}$$

where the second equality holds because volume is not changed by the isometry S (by 10.53), the third equality holds by 10.52 (applied to the positive operator $\sqrt{T^*T}$), and the fourth equality holds by 10.47. ■

The result that we just proved leads to the appearance of determinants in the formula for change of variables in multivariable integration. To describe this, we will again be vague and intuitive.

Throughout this book, almost all the functions we have encountered have been linear. Thus please be aware that the functions f and σ in the material below are not assumed to be linear.

The next definition aims at conveying the idea of the integral; it is not intended as a rigorous definition.

10.55 Definition ***integral,*** $\int_\Omega f$

If $\Omega \subset \mathbf{R}^n$ and f is a real-valued function on Ω, then the ***integral*** of f over Ω, denoted $\int_\Omega f$ or $\int_\Omega f(x)\,dx$, is defined by breaking Ω into pieces small enough that f is almost constant on each piece. On each piece, multiply the (almost constant) value of f by the volume of the piece, then add up these numbers for all the pieces, getting an approximation to the integral that becomes more accurate as Ω is divided into finer pieces.

Actually, Ω in the definition above needs to be a reasonable set (for example, open or measurable) and f needs to be a reasonable function (for example, continuous or measurable), but we will not worry about those technicalities. Also, notice that the x in $\int_\Omega f(x)\,dx$ is a dummy variable and could be replaced with any other symbol.

Now we define the notions of differentiable and derivative. Notice that in this context, the derivative is an operator, not a number as in one-variable calculus. The uniqueness of T in the definition below is left as Exercise 9.

10.56 Definition ***differentiable, derivative,*** $\sigma'(x)$

Suppose Ω is an open subset of $\mathbf{R}^n$ and σ is a function from Ω to $\mathbf{R}^n$. For $x \in \Omega$, the function σ is called ***differentiable*** at x if there exists an operator $T \in \mathcal{L}(\mathbf{R}^n)$ such that

$$\lim_{y \to 0} \frac{\|\sigma(x+y) - \sigma(x) - Ty\|}{\|y\|} = 0.$$

If σ is differentiable at x, then the unique operator $T \in \mathcal{L}(\mathbf{R}^n)$ satisfying the equation above is called the ***derivative*** of σ at x and is denoted by $\sigma'(x)$.

If $n = 1$, then the derivative in the sense of the definition above is the operator on **R** *of multiplication by the derivative in the usual sense of one-variable calculus.*

The idea of the derivative is that for x fixed and $\|y\|$ small,

$$\sigma(x+y) \approx \sigma(x) + \bigl(\sigma'(x)\bigr)(y);$$

because $\sigma'(x) \in \mathcal{L}(\mathbf{R}^n)$, this makes sense.

Suppose Ω is an open subset of $\mathbf{R}^n$ and σ is a function from Ω to $\mathbf{R}^n$. We can write

$$\sigma(x) = \bigl(\sigma_1(x), \dots, \sigma_n(x)\bigr),$$

where each σ_j is a function from Ω to $\mathbf{R}$. The partial derivative of σ_j with respect to the k^{th} coordinate is denoted $D_k\sigma_j$. Evaluating this partial derivative at a point $x \in \Omega$ gives $D_k\sigma_j(x)$. If σ is differentiable at x, then the matrix of $\sigma'(x)$ with respect to the standard basis of $\mathbf{R}^n$ contains $D_k\sigma_j(x)$ in row j, column k (this is left as an exercise). In other words,

10.57
$$\mathcal{M}\bigl(\sigma'(x)\bigr) = \begin{pmatrix} D_1\sigma_1(x) & \dots & D_n\sigma_1(x) \\ \vdots & & \vdots \\ D_1\sigma_n(x) & \dots & D_n\sigma_n(x) \end{pmatrix}.$$

Now we can state the change of variables integration formula. Some additional mild hypotheses are needed for f and σ' (such as continuity or measurability), but we will not worry about them because the proof below is really a pseudoproof that is intended to convey the reason the result is true.

The result below is called a change of variables formula because you can think of $y = \sigma(x)$ as a change of variables, as illustrated by the two examples that follow the proof.

10.58 Change of variables in an integral

Suppose Ω is an open subset of $\mathbf{R}^n$ and $\sigma\colon \Omega \to \mathbf{R}^n$ is differentiable at every point of Ω. If f is a real-valued function defined on $\sigma(\Omega)$, then

$$\int_{\sigma(\Omega)} f(y)\,dy = \int_\Omega f\bigl(\sigma(x)\bigr)\lvert\det\sigma'(x)\rvert\,dx.$$

Proof Let $x \in \Omega$ and let Γ be a small subset of Ω containing x such that f is approximately equal to the constant $f\bigl(\sigma(x)\bigr)$ on the set $\sigma(\Gamma)$.

Adding a fixed vector [such as $\sigma(x)$] to each vector in a set produces another set with the same volume. Thus our approximation for σ near x using the derivative shows that

$$\text{volume}\,\sigma(\Gamma) \approx \text{volume}\big[\big(\sigma'(x)\big)(\Gamma)\big].$$

Using 10.54 applied to the operator $\sigma'(x)$, this becomes

$$\text{volume}\,\sigma(\Gamma) \approx |\det \sigma'(x)|(\text{volume}\,\Gamma).$$

Let $y = \sigma(x)$. Multiply the left side of the equation above by $f(y)$ and the right side by $f\big(\sigma(x)\big)$ [because $y = \sigma(x)$, these two quantities are equal], getting

$$f(y)\,\text{volume}\,\sigma(\Gamma) \approx f\big(\sigma(x)\big)|\det \sigma'(x)|(\text{volume}\,\Gamma).$$

Now break Ω into many small pieces and add the corresponding versions of the equation above, getting the desired result. ■

The key point when making a change of variables is that the factor of $|\det \sigma'(x)|$ must be included when making a substitution $y = f(x)$, as in the right side of 10.58. We finish up by illustrating this point with two important examples.

10.59 Example ***polar coordinates***

Define $\sigma : \mathbf{R}^2 \to \mathbf{R}^2$ by

$$\sigma(r,\theta) = (r\cos\theta, r\sin\theta),$$

where we have used r, θ as the coordinates instead of x_1, x_2 for reasons that will be obvious to everyone familiar with polar coordinates (and will be a mystery to everyone else). For this choice of σ, the matrix of partial derivatives corresponding to 10.57 is

$$\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

as you should verify. The determinant of the matrix above equals r, thus explaining why a factor of r is needed when computing an integral in polar coordinates.

For example, note the extra factor of r in the following familiar formula involving integrating a function f over a disk in $\mathbf{R}^2$:

$$\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} f(x,y)\,dy\,dx = \int_0^{2\pi}\int_0^1 f(r\cos\theta, r\sin\theta)r\,dr\,d\theta.$$

10.60 Example ***spherical coordinates***

Define $\sigma : \mathbf{R}^3 \to \mathbf{R}^3$ by

$$\sigma(\rho, \varphi, \theta) = (\rho \sin\varphi \cos\theta, \rho \sin\varphi \sin\theta, \rho \cos\varphi),$$

where we have used ρ, θ, φ as the coordinates instead of x_1, x_2, x_3 for reasons that will be obvious to everyone familiar with spherical coordinates (and will be a mystery to everyone else). For this choice of σ, the matrix of partial derivatives corresponding to 10.57 is

$$\begin{pmatrix} \sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \\ \cos\varphi & -\rho\sin\varphi & 0 \end{pmatrix},$$

as you should verify. The determinant of the matrix above equals $\rho^2 \sin\varphi$, thus explaining why a factor of $\rho^2 \sin\varphi$ is needed when computing an integral in spherical coordinates.

For example, note the extra factor of $\rho^2 \sin\varphi$ in the following familiar formula involving integrating a function f over a ball in $\mathbf{R}^3$:

$$\int_{-1}^{1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \int_{-\sqrt{1-x^2-y^2}}^{\sqrt{1-x^2-y^2}} f(x, y, z)\, dz\, dy\, dx$$

$$= \int_0^{2\pi} \int_0^{\pi} \int_0^{1} f(\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi)\rho^2 \sin\varphi\, d\rho\, d\varphi\, d\theta.$$

EXERCISES 10.B

1 Suppose V is a real vector space. Suppose $T \in \mathcal{L}(V)$ has no eigenvalues. Prove that $\det T > 0$.

2 Suppose V is a real vector space with even dimension and $T \in \mathcal{L}(V)$. Suppose $\det T < 0$. Prove that T has at least two distinct eigenvalues.

3 Suppose $T \in \mathcal{L}(V)$ and $n = \dim V > 2$. Let $\lambda_1, \dots, \lambda_n$ denote the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space), repeated according to multiplicity.

(a) Find a formula for the coefficient of z^{n-2} in the characteristic polynomial of T in terms of $\lambda_1, \dots, \lambda_n$.

(b) Find a formula for the coefficient of z in the characteristic polynomial of T in terms of $\lambda_1, \dots, \lambda_n$.

4 Suppose $T \in \mathcal{L}(V)$ and $c \in \mathbf{F}$. Prove that $\det(cT) = c^{\dim V} \det T$.

5 Prove or give a counterexample: if $S, T \in \mathcal{L}(V)$, then $\det(S + T) = \det S + \det T$.

6 Suppose A is a block upper-triangular matrix

$$A = \begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_j along the diagonal is a square matrix. Prove that

$$\det A = (\det A_1) \cdots (\det A_m).$$

7 Suppose A is an n-by-n matrix with real entries. Let $S \in \mathcal{L}(\mathbf{C}^n)$ denote the operator on $\mathbf{C}^n$ whose matrix equals A, and let $T \in \mathcal{L}(\mathbf{R}^n)$ denote the operator on $\mathbf{R}^n$ whose matrix equals A. Prove that $\operatorname{trace} S = \operatorname{trace} T$ and $\det S = \det T$.

8 Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Prove that

$$\det T^* = \overline{\det T}.$$

Use this to prove that $|\det T| = \det \sqrt{T^*T}$, giving a different proof than was given in 10.47.

9 Suppose Ω is an open subset of $\mathbf{R}^n$ and σ is a function from Ω to $\mathbf{R}^n$. Suppose $x \in \Omega$ and σ is differentiable at x. Prove that the operator $T \in \mathcal{L}(\mathbf{R}^n)$ satisfying the equation in 10.56 is unique.
[*This exercise shows that the notation $\sigma'(x)$ is justified.*]

10 Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ and $x \in \mathbf{R}^n$. Prove that T is differentiable at x and $T'(x) = T$.

11 Find a suitable hypothesis on σ and then prove 10.57.

12 Let a, b, c be positive numbers. Find the volume of the ellipsoid

$$\left\{(x, y, z) \in \mathbf{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} < 1\right\}$$

by finding a set $\Omega \subset \mathbf{R}^3$ whose volume you know and an operator $T \in \mathcal{L}(\mathbf{R}^3)$ such that $T(\Omega)$ equals the ellipsoid above.

Photo Credits

- page 1: Pierre Louis Dumesnil; 1884 copy by Nils Forsberg/Public domain image from *Wikimedia*.
- page 27: George M. Bergman/Archives of the Mathematisches Forschungsinstitut Oberwolfach.
- page 51: Gottlieb Biermann; photo by A. Wittmann/Public domain image from *Wikimedia*.
- page 117: Mostafa Azizi/Public domain image from *Wikimedia*.
- page 131: Hans-Peter Postel/Public domain image from *Wikimedia*.
- page 163: Public domain image from *Wikimedia*.
- page 203: Public domain image from *Wikimedia*. Original painting is in Tate Britain.
- page 224: Spiked Math.
- page 241: Public domain image from *Wikimedia*.
- page 275: Public domain image from *Wikimedia*. Original fresco is in the Vatican.
- page 295: Public domain image from *Wikimedia*.

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6

Symbol Index

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6

Index

S. Axler, *Linear Algebra Done Right*, Undergraduate Texts in Mathematics,
DOI 10.1007/978-3-319-11080-6