

Linear Algebra Done Right

Sheldon Axler

October 2026

		About this Edition	v
		About the Author	vi
		Preface for Students	vii
		Preface for Instructors	viii
		Acknowledgments	xii
Foundations	1 VECTOR SPACES	SECTION 1A $\mathbb{R}^n$ and $\mathbb{C}^n$	2
		SECTION 1B Definition of Vector Space	11
		SECTION 1C Subspaces	17
	2 FINITE-DIMENSIONAL VECTOR SPACES	SECTION 2A Span and Linear Independence	26
		SECTION 2B Bases	38
		SECTION 2C Dimension	43
	3 LINEAR MAPS	SECTION 3A Vector Space of Linear Maps	50
		SECTION 3B Null Spaces and Ranges	57
		SECTION 3C Matrices	68
		SECTION 3D Invertibility and Isomorphisms	80
		SECTION 3E Products and Quotients of Vector Spaces	95
		SECTION 3F Duality	105
Eigenvalues	4 POLYNOMIALS	CHAPTER 4 Polynomials	121
	5 EIGENVALUES AND EIGENVECTORS	SECTION 5A Invariant Subspaces	133
		SECTION 5B The Minimal Polynomial	144
		SECTION 5C Upper-Triangular Matrices	155
		SECTION 5D Diagonalizable Operators	164
		SECTION 5E Commuting Operators	176

Inner Products

Structure

6 INNER PRODUCT SPACES		SECTION 6A Inner Products and Norms	184
		SECTION 6B Orthonormal Bases	199
		SECTION 6C Orthogonal Complements and Minimization Problems	213
7 OPERATORS ON INNER PRODUCT SPACES		SECTION 7A Self-Adjoint and Normal Operators	230
		SECTION 7B Spectral Theorem	245
		SECTION 7C Positive Operators	252
		SECTION 7D Isometries, Unitary Operators, and Matrix Factorization	259
		SECTION 7E Singular Value Decomposition	271
		SECTION 7F Consequences of Singular Value Decomposition	281
8 OPERATORS ON COMPLEX VECTOR SPACES		SECTION 8A Generalized Eigenvectors and Nilpotent Operators	301
		SECTION 8B Generalized Eigenspace Decomposition	312
		SECTION 8C Consequences of Generalized Eigenspace Decomposition	323
		SECTION 8D Trace: A Connection Between Matrices and Operators	330
9 MULTILINEAR ALGEBRA AND DETERMINANTS		SECTION 9A Bilinear Forms and Quadratic Forms	337
		SECTION 9B Alternating Multilinear Forms	350
		SECTION 9C Determinants	358
		SECTION 9D Tensor Products	373
SOLUTIONS		Preface	388
		Solutions to 1A	389
		Solutions to 1B	392
		Solutions to 1C	396
		Solutions to 2A	403

Solutions to 2B	410
Solutions to 2C	416
Solutions to 3A	425
Solutions to 3B	434
Solutions to 3C	448
Solutions to 3D	456
Solutions to 3E	466
Solutions to 3F	473
Solutions to 4	485
Solutions to 5A	489
Solutions to 5B	503
Solutions to 5C	515
Solutions to 5D	521
Solutions to 5E	531
Solutions to 6A	535
Solutions to 6B	550
Solutions to 6C	560
Solutions to 7A	568
Solutions to 7B	579
Solutions to 7C	588
Solutions to 7D	596
Solutions to 7E	603
Solutions to 7F	610
Solutions to 8A	622
Solutions to 8B	630
Solutions to 8C	638
Solutions to 8D	642
Solutions to 9A	645

SOLUTIONS	Solutions to 9B	648
	Solutions to 9C	650
	Solutions to 9D	658
COLOPHON	Colophon	663

About this Edition

This is a re-typeset edition of *Linear Algebra Done Right*, fourth edition, by Sheldon Axler, © 2024 Sheldon Axler. It follows the open access version of the book dated 13 July 2026. The print version of the book is published by Springer.

The book is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0, https://creativecommons.org/licenses/by-nc/4.0), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as appropriate credit is given to the original author and source, a link to the license is provided, and changes are indicated. This edition is made for noncommercial use only.

This edition changes typography, layout and some notation. The text, the mathematics, the exercises, and the numbering of definitions, results, and examples are Axler's. The notation is changed throughout the text, the exercises and the solutions: $\mathscr{L}(V, W)$ and $\mathscr{L}(V)$ are written $\text{Hom}(V, W)$ and $\text{End}(V)$, the polynomials $\mathscr{P}(\mathbf{F})$ and $\mathscr{P}_m(\mathbf{F})$ are written $\mathbf{F}[X]$ and $\mathbf{F}[X]_{\leq m}$, null spaces and ranges are written with ker and im, and the matrix of an operator is written $\text{Mat}(T)$. The photographs that open the chapters and their captions are not reproduced. The indexes (Symbol Index and Index) and the page references of the original are omitted because the pagination differs. The grouping of the chapters into four parts is an addition of this edition. Errors introduced in the re-typesetting are not the author's.

The solutions to the exercises of all nine chapters are by blanketism (https://github.com/blanketism/Axler-Solutions), from the manual of 5 November 2026; they are not part of Axler's book. They appear in the Solutions chapter at the end of the book, grouped by section, and the alternate proofs of the manual's appendix follow the solution of the exercise they belong to.

About the Author

Sheldon Axler received his undergraduate degree from Princeton University, followed by a PhD in mathematics from the University of California at Berkeley.

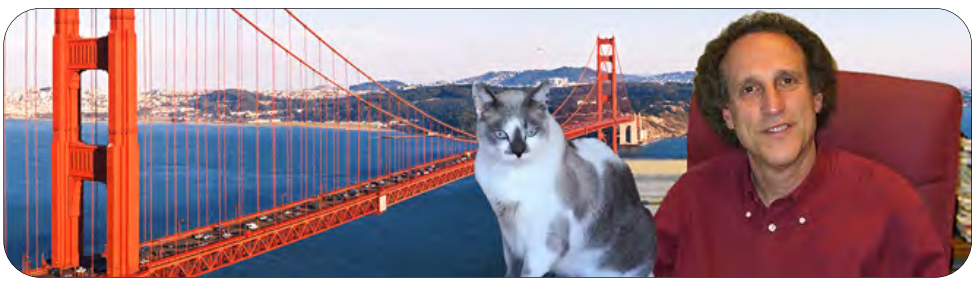
As a postdoctoral Moore Instructor at MIT, Axler received a university-wide teaching award. He was then an assistant professor, associate professor, and professor at Michigan State University, where he received the first J. Sutherland Frame Teaching Award and the Distinguished Faculty Award.

Axler received the Lester R. Ford Award for expository writing from the Mathematical Association of America in 1996, for a paper that eventually expanded into this book. In addition to publishing numerous research papers, he is the author of six mathematics textbooks, ranging from freshman to graduate level. Previous editions of this book have been adopted as a textbook at over 375 universities and colleges and have been translated into three languages.

Axler has served as Editor-in-Chief of the *Mathematical Intelligencer* and Associate Editor of the *American Mathematical Monthly*. He has been a member of the Council of the American Mathematical Society and of the Board of Trustees of the Mathematical Sciences Research Institute. He has also served on the editorial board of Springer's series Undergraduate Texts in Mathematics, Graduate Texts in Mathematics, Universitext, and Springer Monographs in Mathematics.

Axler is a Fellow of the American Mathematical Society and has been a recipient of numerous grants from the National Science Foundation.

Axler joined San Francisco State University as chair of the Mathematics Department in 1997. He served as dean of the College of Science & Engineering from 2002 to 2015, when he returned to a regular faculty appointment as a professor in the Mathematics Department.



The author and his cat Moon.
Carrie Heeter, Bishnu Sarangi

Cover equation: Formula for the nth Fibonacci number. Exercise 21 in Section 5D uses linear algebra to derive this formula.

$$F_n = 1/\sqrt{5}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

Preface for Students

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 and notations are in yellow boxes (and have a title beginning with the word "definition" or "notation") and theorems are in blue boxes (in color versions of the book). Each theorem has an informal descriptive name.

Please check the website below for additional information about the book, including a link to videos that are freely available to accompany the book.

Your suggestions, comments, and corrections are most welcome.

Best wishes for success and enjoyment in learning linear algebra!

Sheldon Axler
San Francisco State University

website: https://linear.axler.net
e-mail: linear@axler.net

Preface for Instructors

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 title of this book deserves an explanation. Most linear algebra textbooks 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 operator 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 5.19) offer more insight. Once determinants have been moved 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. A few examples and exercises involve calculus concepts such as continuity, differentiation, and integration. You can easily skip those examples and exercises if your students have not had calculus. If your students have had calculus, then those examples and exercises can enrich their experience by showing connections between different parts of mathematics.

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: This chapter introduces linear maps. The key result here is the fundamental theorem of linear maps: if T is a linear map on V, then $\dim V = \dim \ker T + \dim \operatorname{im} T$. Quotient spaces and duality are topics in this chapter at a higher level of abstraction than most of the book; these topics can be skipped (except that duality is needed for tensor products in Section 9D).
- 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. The minimal polynomial plays an important role here and later in the book. For example, this chapter gives a characterization of the diagonalizable operators in terms of the minimal polynomial. Section 5E can be skipped if you want to save some time.
- Chapter 6: Inner product spaces are defined in this chapter, and their basic properties are developed along with 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. The pseudoinverse is then introduced as a useful tool when the inverse does not exist. The material on the pseudoinverse can be skipped if you want to save some time.
- Chapter 7: The spectral theorem, which characterizes the linear operators for which there exists an orthonormal basis consisting of eigenvectors, is one of the highlights of this book. The work in earlier chapters pays off here with especially simple proofs. This chapter also deals with positive operators, isometries, unitary operators, matrix factorizations, and especially the singular value decomposition, which leads to the polar decomposition and norms of linear maps.
- Chapter 8: This chapter shows that for each operator on a complex vector space, there is a basis of the vector space consisting of generalized eigenvectors of the operator. Then the generalized eigenspace decomposition describes a linear operator on a complex vector space. The multiplicity of an eigenvalue is defined as the dimension of the corresponding generalized eigenspace. These tools are used to prove that every invertible linear operator on a complex vector space has a square root. Then the chapter gives a proof that every linear operator on a complex vector space can be put into Jordan form. The chapter concludes with an investigation of the trace of operators.
- Chapter 9: This chapter begins by looking at bilinear forms and showing that the vector space of bilinear forms is the direct sum of the subspaces of symmetric bilinear forms and alternating bilinear forms. Then quadratic forms are diagonalized. Moving to multilinear forms, the chapter shows that the subspace of alternating n-linear forms on an n-dimensional vector space has dimension one. This result leads to a clean basis-free definition of the determinant of an operator. For complex vector spaces, the determinant turns out to equal the product of the eigenvalues, with each eigenvalue included in the product as many times as its multiplicity. The chapter concludes with an introduction to tensor products.

This book usually develops linear algebra simultaneously for real and complex vector spaces by letting $\mathbb{F}$ denote either the real or the complex numbers. If you and your students prefer to think of $\mathbb{F}$ as an arbitrary field, then see the comments at the end of Section 1A. 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 all the material in the first seven or eight chapters during a one-semester course may require a rapid pace. If you must reach Chapter 9, then consider skipping the material on quotient spaces in Section 3E, skipping Section 3F on duality (unless you intend to cover tensor products in Section 9D), covering Chapter 4 on polynomials in a half hour, skipping Section 5E on commuting operators, and skipping the subsection in Section 6C on the pseudoinverse.

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 significant time in a typical class.

Some of the exercises are intended to lead curious students into important topics beyond what might usually be included in a basic second course in linear algebra.

The author's top ten

Listed below are the author's ten favorite results in the book, in order of their appearance in the book. Students who leave your course with a good understanding of these crucial results will have an excellent foundation in linear algebra.

- any two bases of a vector space have the same length (2.34)
- fundamental theorem of linear maps (3.21)
- existence of eigenvalues if $\mathbb{F} = \mathbb{C}$ (5.19)
- upper-triangular form always exists if $\mathbb{F} = \mathbb{C}$ (5.47)
- Cauchy–Schwarz inequality (6.14)
- Gram–Schmidt procedure (6.36)
- spectral theorem (7.29 and 7.31)
- singular value decomposition (7.70)
- generalized eigenspace decomposition theorem when $\mathbb{F} = \mathbb{C}$ (8.22)
- dimension of alternating n-linear forms on V is 1 if $\dim V = n$ (9.37)

Major improvements and additions for the fourth edition

- Over 250 new exercises and over 70 new examples.
- Increasing use of the minimal polynomial to provide cleaner proofs of multiple results, including necessary and sufficient conditions for an operator to have an upper-triangular matrix with respect to some basis (see Section 5C), necessary and

sufficient conditions for diagonalizability (see Section 5D), and the real spectral theorem (see Section 7B).

- New section on commuting operators (see Section 5E).
- New subsection on pseudoinverse (see Section 6C).
- New subsections on QR factorization/Cholesky factorization (see Section 7D).
- Singular value decomposition now done for linear maps from an inner product space to another (possibly different) inner product space, rather than only dealing with linear operators from an inner product space to itself (see Section 7E).
- Polar decomposition now proved from singular value decomposition, rather than in the opposite order; this has led to cleaner proofs of both the singular value decomposition (see Section 7E) and the polar decomposition (see Section 7F).
- New subsection on norms of linear maps on finite-dimensional inner product spaces, using the singular value decomposition to avoid even mentioning supremum in the definition of the norm of a linear map (see Section 7F).
- New subsection on approximation by linear maps with lower-dimensional range (see Section 7F).
- New elementary proof of the important result that if T is an operator on a finite-dimensional complex vector space V, then there exists a basis of V consisting of generalized eigenvectors of T (see 8.9).
- New Chapter 9 on multilinear algebra, including bilinear forms, quadratic forms, multilinear forms, and tensor products. Determinants now are defined using a basis-free approach via alternating multilinear forms.
- New formatting to improve the student-friendly appearance of the book. For example, the definition and result boxes now have rounded corners instead of right-angle corners, for a gentler look. The main font size has been reduced from 11 point to 10.5 point.

Please check the website below for additional links and information about the book. Your suggestions, comments, and corrections are most welcome.

Best wishes for teaching a successful linear algebra class!

Sheldon Axler
San Francisco State University

website: https://linear.axler.net
e-mail: linear@axler.net

Contact the author, or Springer if the author is not available, for permission for translations or other commercial reuse of the contents of this book.

Acknowledgments

I owe a huge intellectual debt to all the 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 long ago, then sharpened and improved by many mathematicians in different time periods. Bestowing proper credit on all contributors would be a difficult task that I have not undertaken. In no case should the reader assume that any result presented here represents my original contribution.

Many people helped make this a better book. The three previous editions of this book were used as a textbook at over 375 universities and colleges around the world. I received thousands of suggestions and comments from faculty and students who used the book. Many of those suggestions led to improvements in this edition. The manuscript for this fourth edition was class tested at 30 universities. I am extremely grateful for the useful feedback that I received from faculty and students during this class testing.

The long list of people who should be thanked for their suggestions would fill up many pages. Lists are boring to read. Thus to represent all contributors to this edition, I will mention only Noel Hinton, a graduate student at Australian National University, who sent me more suggestions and corrections for this fourth edition than anyone else. To everyone who contributed suggestions, let me say how truly grateful I am to all of you. Many many thanks!

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 the two terrific mathematics editors at Springer who worked with me on this project—Loretta Bartolini during the first half of my work on the fourth edition, and Elizabeth Loew during the second half of my work on the fourth edition. I am deeply indebted to David Kramer, who did a magnificent job of copyediting and prevented me from making many mistakes.

Extra special thanks to my fantastic partner Carrie Heeter. Her understanding and encouragement enabled me to work intensely on this new edition. Our wonderful cat Moon, whose picture appears on the *About the Author* page, provided sweet breaks throughout the writing process. Moon died suddenly due to a blood clot as this book was being finished. We are grateful for five precious years with him.

Sheldon Axler

Part 1



Foundations

CHAPTER 1	Vector Spaces	2
CHAPTER 2	Finite-Dimensional Vector Spaces	26
CHAPTER 3	Linear Maps	50

1 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 of ordinary space to $\mathbb{R}^n$ and $\mathbb{C}^n$, which we then will generalize to the notion of a vector space. As we will see, a vector space is a set with operations of addition and scalar multiplication that satisfy natural algebraic properties.

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).

1A $\mathbb{R}^n$ and $\mathbb{C}^n$

Complex Numbers

You should already be familiar with basic properties of the set $\mathbb{R}$ 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 by i, that obeys the usual rules of arithmetic. Here are the formal definitions.

Definition 1.1 **Complex numbers, $\mathbb{C}$**

- A *complex number* is an ordered pair (a, b), where $a, b \in \mathbb{R}$, but we will write this as $a + bi$.
- The set of all complex numbers is denoted by $\mathbb{C}$:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}.$$

Definition 1.1 *continued*

- *Addition* and *multiplication* on $\mathbb{C}$ are defined by

$$(a+bi)+(c+di)=(a+c)+(b+d)i,$$
$$(a+bi)(c+di)=(ac-bd)+(ad+bc)i;$$

here $a, b, c, d \in \mathbb{R}$.

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

To motivate the definition of complex multiplication given above, pretend that we knew that $i^2=-1$ and then use the usual rules of arithmetic to derive the formula above for the product of two complex numbers. Then use that formula to verify that we indeed have

$$i^2=-1.$$

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

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). The next example illustrates this procedure.

Example 1.2 **Complex arithmetic**

The product $(2+3i)(4+5i)$ can be evaluated by applying the distributive and commutative properties from 1.3:

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

Our first result states that complex addition and complex multiplication have the familiar properties that we expect.

Result 1.3 **Properties of complex arithmetic**

commutativity
$\alpha+\beta=\beta+\alpha$ and $\alpha\beta=\beta\alpha$ for all $\alpha, \beta \in \mathbb{C}$.

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

identities
$\lambda+0=\lambda$ and $\lambda 1=\lambda$ for all $\lambda \in \mathbb{C}$.

additive inverse
For every $\alpha \in \mathbb{C}$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha+\beta=0$.

Result 1.3
continued

multiplicative inverse
For every $\alpha \in \mathbb{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha\beta = 1$.

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

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.

Example 1.4 **Commutativity of complex multiplication**

To show that $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbb{C}$, suppose

$$\alpha = a + bi \quad \text{and} \quad \beta = c + di,$$

where $a, b, c, d \in \mathbb{R}$. 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 $\alpha\beta = \beta\alpha$.

Next, we define the additive and multiplicative inverses of complex numbers, and then use those inverses to define subtraction and division operations with complex numbers.

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

Suppose $\alpha, \beta \in \mathbb{C}$.

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

Definition 1.5 *continued*

- For $\alpha \neq 0$, *division* by α 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.

Notation 1.6

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

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

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

Elements of $\mathbb{F}$ are called *scalars*. The word "scalar" (which is just 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 $\alpha \in \mathbb{F}$ and m a positive integer, we define α^m to denote the product of α with itself m times:

$$\alpha^m = \underbrace{\alpha \cdots \alpha}_{m \text{ times}}.$$

This definition implies that

$$(\alpha^m)^n = \alpha^{mn} \quad \text{and} \quad (\alpha\beta)^m = \alpha^m \beta^m$$

for all $\alpha, \beta \in \mathbb{F}$ and all positive integers m, n.

Lists

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

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

- The set $\mathbb{R}^2$, which you can think of as a plane, is the set of all ordered pairs of real numbers:

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}.$$

- The set $\mathbb{R}^3$, which you can think of as ordinary space, is the set of all ordered triples of real numbers:

$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}.$$

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

List, length — Definition 1.8

- 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 objects).
- Two lists are equal if and only if they have the same length and the same elements in the same order.

Lists are often written as elements separated by commas and surrounded by parentheses. Thus a list of length two is an ordered pair that might be written as (a, b). A list of length three is an ordered triple that might be written as (x, y, z). A list of length n might look like this:

$$(z_1, \dots, z_n).$$

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 $(x_1, x_2, \dots)$, 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 finite sets in two ways: in lists, order matters and repetitions have meaning; in sets, order and repetitions are irrelevant.

Example 1.9 **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 $\{4\}$.

$\mathbb{F}^n$

To define the higher-dimensional analogues of $\mathbb{R}^2$ and $\mathbb{R}^3$, we will simply replace $\mathbb{R}$ with $\mathbb{F}$ (which equals $\mathbb{R}$ or $\mathbb{C}$) and replace the 2 or 3 with an arbitrary positive integer.

n — Notation 1.10

Fix a positive integer n for the rest of this chapter.

$\mathbb{F}^n$, coordinate — Definition 1.11

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

$$\mathbb{F}^n = \{(x_1, \dots, x_n) : x_k \in \mathbb{F} \text{ for } k = 1, \dots, n\}.$$

For $(x_1, \dots, x_n) \in \mathbb{F}^n$ and $k \in \{1, \dots, n\}$, we say that x_k is the kth *coordinate* of

Definition 1.11 *continued*

$(x_1, \dots, x_n)$.

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

Example 1.12 $\mathbb{C}^4$

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

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

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

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

Addition in $\mathbb{F}^n$ Definition 1.13

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

Commutativity of addition in $\mathbb{F}^n$ Result 1.14

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

Proof. Suppose $x = (x_1, \dots, x_n) \in \mathbb{F}^n$ and $y = (y_1, \dots, y_n) \in \mathbb{F}^n$. 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 $\mathbb{F}^n$ and the third equality holds because of the usual commutativity of addition in $\mathbb{F}$. □

The symbol ■ means "end of proof".

If a single letter is used to denote an element of $\mathbb{F}^n$, then the same letter with appropriate subscripts is often used when coordinates must be displayed. For example, if $x \in \mathbb{F}^n$, 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.

Notation 1.15 0

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

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

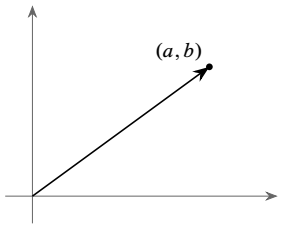
Here we are using the symbol 0 in two different ways—on the left side of the equation above, the symbol 0 denotes a list of length n, which is an element of $\mathbb{F}^n$, whereas on the right side, each 0 denotes a number. This potentially confusing practice actually causes no problems because the context should always make clear which 0 is intended.

Example 1.16 **Context determines which 0 is intended**

Consider the statement that 0 is an additive identity for $\mathbb{F}^n$:

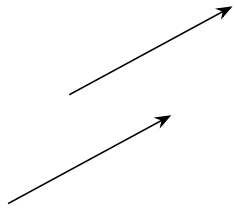
$$x + 0 = x \quad \text{for all } x \in \mathbb{F}^n.$$

Here the 0 above is the list defined in 1.15, not the number 0, because we have not defined the sum of an element of $\mathbb{F}^n$ (namely, x) and the number 0.



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

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



A vector.

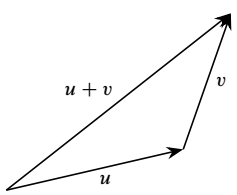
When we think of vectors in $\mathbb{R}^2$ as arrows, we can move an arrow parallel to itself (not changing its length or direction) and still think of it as the same vector. With that viewpoint, you will often gain better understanding by dispensing with the coordinate axes and the explicit coordinates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

Whenever we use pictures in $\mathbb{R}^2$ 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 $\mathbb{R}^2$.

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

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

Recall that we defined the sum of two elements of $\mathbb{F}^n$ to be the element of $\mathbb{F}^n$ obtained by adding corresponding coordinates; see 1.13. As we will now see, addition has a simple geometric interpretation in the special case of $\mathbb{R}^2$.



The sum of two vectors.

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

initial point of u and whose end point equals the end point of the vector v, as shown here.

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

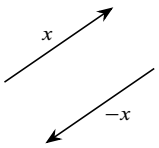
Additive inverse in $\mathbb{F}^n$, $-x$ Definition 1.17

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

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

Thus if $x = (x_1, \ldots, x_n)$, then $-x = (-x_1, \ldots, -x_n)$.

The additive inverse of a vector in $\mathbb{R}^2$ is the vector with the same length but pointing in the opposite direction. The figure here illustrates this way of thinking about the additive inverse in $\mathbb{R}^2$. As you can see, the vector labeled $-x$ has the same length as the vector labeled x but points in the opposite direction.



A vector and its additive inverse.

Having dealt with addition in $\mathbb{F}^n$, we now turn to multiplication. We could define a multiplication in $\mathbb{F}^n$ in a similar fashion, starting with two elements of $\mathbb{F}^n$ and getting another element of $\mathbb{F}^n$ 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 $\mathbb{F}^n$ by an element of $\mathbb{F}$.

Scalar multiplication in $\mathbb{F}^n$ Definition 1.18

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

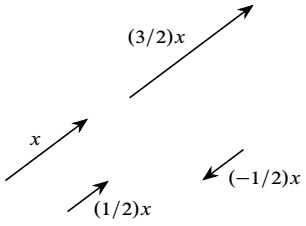
$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n);$$

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

Scalar multiplication in $\mathbb{F}^n$ multiplies together a scalar and a vector, getting a vector. In contrast, the dot product in $\mathbb{R}^2$ or $\mathbb{R}^3$ multiplies together two vectors and gets a scalar. Generalizations of the dot product will become important in Chapter 6.

Scalar multiplication has a nice geometric interpretation in $\mathbb{R}^2$. If $\lambda > 0$ and $x \in \mathbb{R}^2$, 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 $\lambda < 0$ and $x \in \mathbb{R}^2$, 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 properties listed in 1.3. Thus $\mathbb{R}$ and $\mathbb{C}$ 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 deal with fields other than $\mathbb{R}$ and $\mathbb{C}$. However, many of the definitions, theorems, and proofs in linear algebra that work for the fields $\mathbb{R}$ and $\mathbb{C}$ also work without change for arbitrary fields. If you prefer to do so, throughout much of this book (except for Chapters 6 and 7, which deal with inner product spaces) you can think of $\mathbb{F}$ as denoting an arbitrary field instead of $\mathbb{R}$ or $\mathbb{C}$. For results (except in the inner product chapters) that have as a hypothesis that $\mathbb{F}$ is $\mathbb{C}$, you can probably replace that hypothesis with the hypothesis that $\mathbb{F}$ is an algebraically closed field, which means that every nonconstant polynomial with coefficients in $\mathbb{F}$ has a zero. A few results, such as Exercise 13 in Section 1C, require the hypothesis on $\mathbb{F}$ that $1 + 1 \neq 0$.

Exercises

1A.1 Show that $\alpha + \beta = \beta + \alpha$ for all $\alpha, \beta \in \mathbb{C}$.

1A.2 Show that $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ for all $\alpha, \beta, \lambda \in \mathbb{C}$.

1A.3 Show that $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbb{C}$.

1A.4 Show that $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbb{C}$.

1A.5 Show that for every $\alpha \in \mathbb{C}$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha + \beta = 0$.

1A.6 Show that for every $\alpha \in \mathbb{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha\beta = 1$.

1A.7 Show that

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

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

1A.8 Find two distinct square roots of i.

1A.9 Find $x \in \mathbb{R}^4$ such that

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

1A.10 Explain why there does not exist $\lambda \in \mathbb{C}$ such that

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

1A.11 Show that $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbb{F}^n$.

1A.12 Show that $(ab)x = a(bx)$ for all $x \in \mathbb{F}^n$ and all $a, b \in \mathbb{F}$.

1A.13 Show that $1x = x$ for all $x \in \mathbb{F}^n$.

1A.14 Show that $\lambda(x + y) = \lambda x + \lambda y$ for all $\lambda \in \mathbb{F}$ and all $x, y \in \mathbb{F}^n$.

1A.15 Show that $(a + b)x = ax + bx$ for all $a, b \in \mathbb{F}$ and all $x \in \mathbb{F}^n$.

"Can you do addition?" the White Queen asked. "What's one and one and one and one and one and one and one and one and one and one?"
"I don't know," said Alice. "I lost count."

—Through the Looking Glass, Lewis Carroll

1B Definition of Vector Space

The motivation for the definition of a vector space comes from properties of addition and scalar multiplication in $\mathbb{F}^n$: 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.

Addition, scalar multiplication — **Definition 1.19**

- 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 $\lambda \in \mathbb{F}$ and each $v \in V$.

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

Vector space — **Definition 1.20**

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 for all

Definition 1.20 *continued*

$a, b \in \mathbb{F}$.

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 $a, b \in \mathbb{F}$ and all $u, v \in V$.

The following geometric language sometimes aids our intuition.

Vector, point — Definition 1.21

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

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

Real vector space, complex vector space — Definition 1.22

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

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

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

The simplest vector space is $\{0\}$, which contains only one point.

Example 1.23 $\mathbb{F}^\infty$

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

$$\mathbb{F}^\infty = \{(x_1, x_2, \dots) : x_k \in \mathbb{F} \text{ for } k = 1, 2, \dots\}.$$

Addition and scalar multiplication on $\mathbb{F}^\infty$ 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, $\mathbb{F}^\infty$ becomes a vector space over $\mathbb{F}$, 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.

Notation 1.24

$\mathbb{F}^S$

- If S is a set, then $\mathbb{F}^S$ denotes the set of functions from S to $\mathbb{F}$.
- For $f, g \in \mathbb{F}^S$, the *sum* $f + g \in \mathbb{F}^S$ is the function defined by

$$(f + g)(x) = f(x) + g(x)$$

  for all $x \in S$.
- For $\lambda \in \mathbb{F}$ and $f \in \mathbb{F}^S$, the *product* $\lambda f \in \mathbb{F}^S$ 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 $\mathbb{F} = \mathbb{R}$, then $\mathbb{R}^{[0,1]}$ is the set of real-valued functions on the interval $[0, 1]$.

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

Example 1.25 **$\mathbb{F}^S$ is a vector space**

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

$$0(x) = 0$$

  for all $x \in S$.
- For $f \in \mathbb{F}^S$, the additive inverse of f is the function $-f: S \to \mathbb{F}$ defined by

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

  for all $x \in S$.

The vector space $\mathbb{F}^n$ is a special case of the vector space $\mathbb{F}^S$ because each $(x_1, \ldots, x_n) \in \mathbb{F}^n$ can be thought of as a function x from the set $\{1, 2, \ldots, n\}$ to $\mathbb{F}$ by writing $x(k)$ instead of x_k for the kth coordinate of $(x_1, \ldots, x_n)$. In other words, we can think of $\mathbb{F}^n$ as $\mathbb{F}^{\{1,2,\ldots,n\}}$. Similarly, we can think of $\mathbb{F}^\infty$ as $\mathbb{F}^{\{1,2,\ldots\}}$.

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

The elements of the vector space $\mathbb{R}^{[0,1]}$ 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.

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

Unique additive identity — Result 1.26

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.

Unique additive inverse — Result 1.27

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.

$-v$, $w - v$ — Notation 1.28

Let $v, w \in V$. Then

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

Almost all 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.

V — Notation 1.29

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

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

The number 0 times a vector **Result 1.30**
$0v = 0$ for every $v \in V$.

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. □

The result in 1.30 involves the additive identity of V and scalar multiplication. The only part of the definition of a vector space that connects vector addition and scalar multiplication is the distributive property. Thus the distributive property must be used in the proof of 1.30.

In the next result, 0 denotes the additive identity of V. Although their proofs are similar, 1.30 and 1.31 are not identical. More precisely, 1.30 states that the product of the scalar 0 and any vector equals the vector 0, whereas 1.31 states that the product of any scalar and the vector 0 equals the vector 0.

A number times the vector 0 **Result 1.31**
$a0 = 0$ for every $a \in \mathbb{F}$.

Proof. For $a \in \mathbb{F}$, 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.

The number -1 times a vector **Result 1.32**
$(-1)v = -v$ 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

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

1B.2 Suppose $a \in \mathbb{F}$, $v \in V$, and $av = 0$. Prove that $a = 0$ or $v = 0$.

1B.3 Suppose $v, w \in V$. Explain why there exists a unique $x \in V$ such that $v + 3x = w$.

1B.4 The empty set is not a vector space. The empty set fails to satisfy only one of the requirements listed in the definition of a vector space (1.20). Which one?

1B.5 Show that in the definition of a vector space (1.20), 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.

1B.6 Let ∞ and $-\infty$ denote two distinct objects, neither of which is in $\mathbb{R}$. Define an addition and scalar multiplication on $\mathbb{R} \cup \{\infty, -\infty\}$ as you could guess from the notation. Specifically, the sum and product of two real numbers is as usual, and for $t \in \mathbb{R}$ define

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

and

$$\begin{gathered} t + \infty = \infty + t = \infty + \infty = \infty, \\ t + (-\infty) = (-\infty) + t = (-\infty) + (-\infty) = -\infty, \\ \infty + (-\infty) = (-\infty) + \infty = 0. \end{gathered}$$

With these operations of addition and scalar multiplication, is $\mathbb{R} \cup \{\infty, -\infty\}$ a vector space over $\mathbb{R}$? Explain.

1B.7 Suppose S is a nonempty set. Let V^S denote the set of functions from S to V. Define a natural addition and scalar multiplication on V^S, and show that V^S is a vector space with these definitions.

1B.8 Suppose V is a real vector space.

- The *complexification* of V, denoted by $V_{\mathbb{C}}$, equals $V \times V$. An element of $V_{\mathbb{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we write this as $u + iv$.
- Addition on $V_{\mathbb{C}}$ is defined by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

  for all $u_1, v_1, u_2, v_2 \in V$.
- Complex scalar multiplication on $V_{\mathbb{C}}$ is defined by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

  for all $a, b \in \mathbb{R}$ and all $u, v \in V$.

Prove that with the definitions of addition and scalar multiplication as above, $V_{\mathbb{C}}$ is a complex vector space.

Think of V as a subset of $V_{\mathbb{C}}$ by identifying $u \in V$ with $u + i0$. The construction of $V_{\mathbb{C}}$ from V can then be thought of as generalizing the construction of $\mathbb{C}^n$ from $\mathbb{R}^n$.

1C Subspaces

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

Some people use the terminology *linear subspace*, which means the same as subspace.

Subspace — **Definition 1.33**

A subset U of V is called a *subspace* of V if U is also a vector space with the same additive identity, addition, and scalar multiplication as on V.

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

Conditions for a subspace — **Result 1.34**

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

The additive identity condition above could be replaced with the condition that U is nonempty (because then taking $u \in U$ and multiplying it by 0 would imply that $0 \in U$). However, if a subset U of V is indeed a subspace, then usually the quickest 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 ensures that the additive identity of V is in U. The second condition 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 1.32] 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 assertions in the next example.

Example 1.35 **Subspaces**

(a) If $b \in \mathbb{F}$, then
$$\{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 : x_3 = 5x_4 + b\}$$
is a subspace of $\mathbb{F}^4$ if and only if $b = 0$.

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

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

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

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

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

The set $\{0\}$ 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.

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

Sums of Subspaces

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.

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

Sum of subspaces Definition 1.36

Suppose $V_1, \dots, V_m$ are subspaces of V. The *sum* of $V_1, \dots, V_m$, denoted by $V_1 + \cdots + V_m$, is the set of all possible sums of elements of $V_1, \dots, V_m$. More precisely,

$$V_1 + \cdots + V_m = \{v_1 + \cdots + v_m : v_1 \in V_1, \dots, v_m \in V_m\}.$$

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

Example 1.37 **A sum of subspaces of $\mathbb{F}^3$**

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

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

Then

$$U + W = \{(x, y, 0) \in \mathbb{F}^3 : x, y \in \mathbb{F}\},$$

as you should verify.

Example 1.38 **A sum of subspaces of $\mathbb{F}^4$**

Suppose

$$U = \{(x, x, y, y) \in \mathbb{F}^4 : x, y \in \mathbb{F}\} \quad \text{and} \quad W = \{(x, x, x, y) \in \mathbb{F}^4 : x, y \in \mathbb{F}\}.$$

Using words rather than symbols, we could say that U is the set of elements of $\mathbb{F}^4$ whose first two coordinates equal each other and whose third and fourth coordinates equal each other. Similarly, W is the set of elements of $\mathbb{F}^4$ whose first three coordinates equal each other.

To find a description of $U + W$, consider a typical element (a, a, b, b) of U and a typical element (c, c, c, d) of W, where $a, b, c, d \in \mathbb{F}$. We have

$$(a, a, b, b) + (c, c, c, d) = (a + c, a + c, b + c, b + d),$$

which shows that every element of $U + W$ has its first two coordinates equal to each other. Thus

$$U + W \subseteq \{(x, x, y, z) \in \mathbb{F}^4 : x, y, z \in \mathbb{F}\}. \tag{1.39}$$

To prove the inclusion in the other direction, suppose $x, y, z \in \mathbb{F}$. Then

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

where the first vector on the right is in U and the second vector on the right is in W. Thus $(x, x, y, z) \in U + W$, showing that the inclusion 1.39 also holds in the

Example 1.38
continued

opposite direction. Hence

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

which shows that $U + W$ is the set of elements of $\mathbb{F}^4$ whose first two coordinates equal each other.

The next result states that the sum of subspaces is a subspace, and is in fact the smallest subspace containing all the summands (which means that every subspace containing all the summands also contains the sum).

Sum of subspaces is the smallest containing subspace Result 1.40

Suppose $V_1, \dots, V_m$ are subspaces of V. Then $V_1 + \cdots + V_m$ is the smallest subspace of V containing $V_1, \dots, V_m$.

Proof. The reader can verify that $V_1 + \cdots + V_m$ contains the additive identity 0 and is closed under addition and scalar multiplication. Thus 1.34 implies that $V_1 + \cdots + V_m$ is a subspace of V.

The subspaces $V_1, \dots, V_m$ are all contained in $V_1 + \cdots + V_m$ (to see this, consider sums $v_1 + \cdots + v_m$ where all except one of the v_k's are 0). Conversely, every subspace of V containing $V_1, \dots, V_m$ contains $V_1 + \cdots + V_m$ (because subspaces must contain all finite sums of their elements). Thus $V_1 + \cdots + V_m$ is the smallest subspace of V containing $V_1, \dots, V_m$. □

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 $V_1, \dots, V_m$ are subspaces of V. Every element of $V_1 + \cdots + V_m$ can be written in the form

$$v_1 + \cdots + v_m,$$

where each $v_k \in V_k$. Of special interest are cases in which each vector in $V_1 + \cdots + V_m$ can be represented in the form above in only one way. This situation is so important that it gets a special name (direct sum) and a special symbol ($\oplus$).

Direct sum, $\oplus$ Definition 1.41

Suppose $V_1, \dots, V_m$ are subspaces of V.

- The sum $V_1 + \cdots + V_m$ is called a *direct sum* if each element of $V_1 + \cdots + V_m$ can be written in only one way as a sum $v_1 + \cdots + v_m$, where each $v_k \in V_k$.
- If $V_1 + \cdots + V_m$ is a direct sum, then $V_1 \oplus \cdots \oplus V_m$ denotes $V_1 + \cdots + V_m$, with the $\oplus$ notation serving as an indication that this is a direct sum.

Example 1.42 **A direct sum of two subspaces**

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

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

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

To produce $\oplus$ in TEX, type `\oplus`.

Example 1.43 **A direct sum of multiple subspaces**

Suppose V_k is the subspace of $\mathbb{F}^n$ of those vectors whose coordinates are all 0, except possibly in the kth slot; for example, $V_2 = \{(0, x, 0, \dots, 0) \in \mathbb{F}^n : x \in \mathbb{F}\}$. Then

$$\mathbb{F}^n = V_1 \oplus \cdots \oplus V_n,$$

as you should verify.

Sometimes nonexamples add to our understanding as much as examples.

Example 1.44 **A sum that is not a direct sum**

Suppose

$$V_1 = \{(x, y, 0) \in \mathbb{F}^3 : x, y \in \mathbb{F}\},$$
$$V_2 = \{(0, 0, z) \in \mathbb{F}^3 : z \in \mathbb{F}\},$$
$$V_3 = \{(0, y, y) \in \mathbb{F}^3 : y \in \mathbb{F}\}.$$

Then $\mathbb{F}^3 = V_1 + V_2 + V_3$ because every vector $(x, y, z) \in \mathbb{F}^3$ 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 V_1, the second vector is in V_2, and the third vector is in V_3.

However, $\mathbb{F}^3$ does not equal the direct sum of V_1, V_2, V_3, because the vector $(0, 0, 0)$ can be written in more than one way as a sum $v_1 + v_2 + v_3$, with each $v_k \in V_k$. 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 V_1, the second vector is in V_2, and the third vector is in V_3. Thus the sum $V_1 + V_2 + V_3$ is not a direct sum.

The definition of direct sum requires every vector in the sum to 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 only need to consider whether 0 can be uniquely written as an appropriate sum.

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

Condition for a direct sum **Result 1.45**

Suppose $V_1, \dots, V_m$ are subspaces of V. Then $V_1 + \cdots + V_m$ is a direct sum if and only if the only way to write 0 as a sum $v_1 + \cdots + v_m$, where each $v_k \in V_k$, is by taking each v_k equal to 0.

Proof. First suppose $V_1 + \cdots + V_m$ is a direct sum. Then the definition of direct sum implies that the only way to write 0 as a sum $v_1 + \cdots + v_m$, where each $v_k \in V_k$, is by taking each v_k equal to 0.

Now suppose that the only way to write 0 as a sum $v_1 + \cdots + v_m$, where each $v_k \in V_k$, is by taking each v_k equal to 0. To show that $V_1 + \cdots + V_m$ is a direct sum, let $v \in V_1 + \cdots + V_m$. We can write

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

for some $v_1 \in V_1, \dots, v_m \in V_m$. To show that this representation is unique, suppose we also have

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

where $u_1 \in V_1, \dots, u_m \in V_m$. Subtracting these two equations, we have

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

The symbol $\iff$ used below means "if and only if"; this symbol could also be read to mean "is equivalent to".

Because $v_1 - u_1 \in V_1, \dots, v_m - u_m \in V_m$, the equation above implies that each $v_k - u_k$ equals 0. Thus $v_1 = u_1, \dots, v_m = u_m$, as desired. □

The next result gives a simple condition for testing whether a sum of two subspaces is a direct sum.

Direct sum of two subspaces **Result 1.46**

Suppose U and W are subspaces of V. Then

$$U + W \text{ is a direct sum} \iff U \cap W = \{0\}.$$

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 only need to show that $u = w = 0$ (by 1.45). 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. □

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 1.44. In that nonexample of a direct sum, we have $V_1 \cap V_2 = V_1 \cap V_3 = V_2 \cap V_3 = \{0\}$.

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 equal $\{0\}$.

Exercises

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

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

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

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

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

1C.2 Verify all assertions about subspaces in Example 1.35.

1C.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 $\mathbb{R}^{(-4,4)}$.

1C.4 Suppose $b \in \mathbb{R}$. 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 $\mathbb{R}^{[0,1]}$ if and only if $b = 0$.

1C.5 Is $\mathbb{R}^2$ a subspace of the complex vector space $\mathbb{C}^2$?

1C.6

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

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

1C.7 Prove or give a counterexample: If U is a nonempty subset of $\mathbb{R}^2$ such that U is closed under addition and under taking additive inverses (meaning $-u \in U$ whenever $u \in U$), then U is a subspace of $\mathbb{R}^2$.

1C.8 Give an example of a nonempty subset U of $\mathbb{R}^2$ such that U is closed under scalar multiplication, but U is not a subspace of $\mathbb{R}^2$.

1C.9 A function $f : \mathbb{R} \to \mathbb{R}$ is called *periodic* if there exists a positive number p such that $f(x) = f(x + p)$ for all $x \in \mathbb{R}$. Is the set of periodic functions from $\mathbb{R}$ to $\mathbb{R}$ a subspace of $\mathbb{R}^{\mathbb{R}}$? Explain.

1C.10 Suppose V_1 and V_2 are subspaces of V. Prove that the intersection $V_1 \cap V_2$ is a subspace of V.

1C.11 Prove that the intersection of every collection of subspaces of V is a subspace of V.

1C.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.

1C.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 Exercise 12, possibly because this exercise is not true if we replace $\mathbb{F}$ with a field containing only two elements.

1C.14 Suppose

$$U = \{(x, -x, 2x) \in \mathbb{F}^3 : x \in \mathbb{F}\} \quad \text{and} \quad W = \{(x, x, 2x) \in \mathbb{F}^3 : x \in \mathbb{F}\}.$$

Describe $U + W$ using symbols, and also give a description of $U + W$ that uses no symbols.

1C.15 Suppose U is a subspace of V. What is $U + U$?

1C.16 Is the operation of addition on the subspaces of V commutative? In other words, if U and W are subspaces of V, is $U + W = W + U$?

1C.17 Is the operation of addition on the subspaces of V associative? In other words, if V_1, V_2, V_3 are subspaces of V, is

$$(V_1 + V_2) + V_3 = V_1 + (V_2 + V_3)?$$

1C.18 Does the operation of addition on the subspaces of V have an additive identity? Which subspaces have additive inverses?

1C.19 Prove or give a counterexample: If V_1, V_2, U are subspaces of V such that

$$V_1 + U = V_2 + U,$$

then $V_1 = V_2$.

1C.20 Suppose

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

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

1C.21 Suppose

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

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

1C.22 Suppose

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

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

1C.23 Prove or give a counterexample: If V_1, V_2, U are subspaces of V such that

$$V = V_1 \oplus U \quad \text{and} \quad V = V_2 \oplus U,$$

then $V_1 = V_2$.

Hint: When trying to discover whether a conjecture in linear algebra is true or false, it is often useful to start by experimenting in $\mathbb{F}^2$.

1C.24 A function $f : \mathbb{R} \to \mathbb{R}$ is called *even* if

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

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

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

for all $x \in \mathbb{R}$. Let V_e denote the set of real-valued even functions on $\mathbb{R}$ and let V_o denote the set of real-valued odd functions on $\mathbb{R}$. Show that $\mathbb{R}^{\mathbb{R}} = V_e \oplus V_o$.

Finite-Dimensional Vector Spaces



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.

We begin this chapter by considering linear combinations of lists of vectors. This leads us to the crucial concept of linear independence. The linear dependence lemma will become one of our most useful tools.

A list of vectors in a vector space that is small enough to be linearly independent and big enough so the linear combinations of the list fill up the vector space is called a basis of the vector space. We will see that every basis of a vector space has the same length, which will allow us to define the dimension of a vector space.

This chapter ends with a formula for the dimension of the sum of two subspaces.

Standing assumptions for this chapter — Definition

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

2A Span and Linear Independence

We have been writing lists of numbers surrounded by parentheses, and we will continue to do so for elements of $\mathbb{F}^n$; for example, $(2, -7, 8) \in \mathbb{F}^3$. However, now we need to consider lists of vectors (which may be elements of $\mathbb{F}^n$ 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 two of vectors in $\mathbb{R}^3$.

List of vectors — Notation 2.1

We will usually write lists of vectors without surrounding parentheses.

Linear Combinations and Span

A sum of scalar multiples of the vectors in a list is called a linear combination of the list. Here is the formal definition.

Linear combination — Definition 2.2

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 $a_1, \dots, a_m \in \mathbb{F}$.

Example 2.3 **Linear combinations in $\mathbb{R}^3$**

- $(17, -4, 2)$ is a linear combination of $(2, 1, -3), (1, -2, 4)$, which is a list of length two of vectors in $\mathbb{R}^3$, because

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

- $(17, -4, 5)$ is not a linear combination of $(2, 1, -3), (1, -2, 4)$, which is a list of length two of vectors in $\mathbb{R}^3$, because there do not exist numbers $a_1, a_2 \in \mathbb{R}$ such that

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

  In other words, the system of equations

$$17 = 2a_1 + a_2$$
$$-4 = a_1 - 2a_2$$
$$5 = -3a_1 + 4a_2$$

  has no solutions (as you should verify).

Span — Definition 2.4

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

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

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

Example 2.5 **Span**

The previous example shows that in $\mathbb{F}^3$,

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

Result 2.6

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

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

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

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

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

$$(a_1v_1 + \cdots + a_mv_m) + (c_1v_1 + \cdots + c_mv_m) = (a_1 + c_1)v_1 + \cdots + (a_m + c_m)v_m.$$

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

$$\lambda(a_1v_1 + \cdots + a_mv_m) = \lambda a_1v_1 + \cdots + \lambda a_mv_m.$$

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

Each v_k is a linear combination of $v_1, \dots, v_m$ (to show this, set $a_k = 1$ and let the other a's in 2.2 equal 0). Thus $\operatorname{span}(v_1, \dots, v_m)$ contains each v_k. Conversely, because subspaces are closed under scalar multiplication and addition, every subspace of V that contains each v_k contains $\operatorname{span}(v_1, \dots, v_m)$. Thus $\operatorname{span}(v_1, \dots, v_m)$ is the smallest subspace of V containing all the vectors $v_1, \dots, v_m$. □

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

Definition 2.7

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

Example 2.8 **A list that spans $\mathbb{F}^n$**

Suppose n is a positive integer. We want to show that

$$(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$$

spans $\mathbb{F}^n$. Here the kth vector in the list above has 1 in the kth slot and 0 in all other slots.

Example 2.8
continued

Suppose $(x_1, \dots, x_n) \in \mathbb{F}^n$. Then

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

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

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

Finite-dimensional vector space Definition 2.9

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

Example 2.8 above shows that $\mathbb{F}^n$ is a finite-dimensional vector space for every positive integer n.

Recall that by definition every list has finite length.

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

Polynomial, $\mathbb{F}[X]$ Definition 2.10

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

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

  for all $z \in \mathbb{F}$.
- $\mathbb{F}[X]$ is the set of all polynomials with coefficients in $\mathbb{F}$.

With the usual operations of addition and scalar multiplication, $\mathbb{F}[X]$ is a vector space over $\mathbb{F}$, as you should verify. Hence $\mathbb{F}[X]$ is a subspace of $\mathbb{F}^{\mathbb{F}}$, the vector space of functions from $\mathbb{F}$ to $\mathbb{F}$.

If a polynomial (thought of as a function from $\mathbb{F}$ to $\mathbb{F}$) 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 $\mathbb{F}$ and hence has all zero coefficients (if you are unfamiliar with this fact, just believe it for now; we will prove it later—see 4.8). **Conclusion:** the coefficients of a polynomial are uniquely determined by the polynomial. Thus the next definition uniquely defines the degree of a polynomial.

Degree of a polynomial, $\deg p$ Definition 2.11

- A polynomial $p \in \mathbb{F}[X]$ is said to have *degree* m if there exist scalars $a_0, a_1, \dots, a_m \in \mathbb{F}$ with $a_m \neq 0$ such that for every $z \in \mathbb{F}$, we have

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

Definition 2.11 *continued*

- The polynomial that is identically 0 is said to have degree $-\infty$.
- The degree of a polynomial p is denoted by $\deg p$.

In the next definition, we use the convention that $-\infty < m$, which means that the polynomial 0 is in $\mathbb{F}[X]_{\le m}$.

Notation 2.12

$\mathbb{F}[X]_{\le m}$
For m a nonnegative integer, $\mathbb{F}[X]_{\le m}$ denotes the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m.

If m is a nonnegative integer, then $\mathbb{F}[X]_{\le m} = \operatorname{span}(1, z, \ldots, z^m)$ [here we slightly abuse notation by letting z^k denote a function]. Thus $\mathbb{F}[X]_{\le m}$ is a finite-dimensional vector space for each nonnegative integer m.

Definition 2.13

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

Example 2.14 $\mathbb{F}[X]$ **is infinite-dimensional**

Consider any list of elements of $\mathbb{F}[X]$. 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 $\mathbb{F}[X]$. Thus $\mathbb{F}[X]$ is infinite-dimensional.

Linear Independence

Suppose $v_1, \ldots, v_m \in V$ and $v \in \operatorname{span}(v_1, \ldots, v_m)$. By the definition of span, there exist $a_1, \ldots, a_m \in \mathbb{F}$ 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 $c_1, \ldots, c_m$ 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 $(v_1, \ldots, v_m)$. If the only way to do this is by using 0 for all the scalars in the linear combination, then each $a_k - c_k$ equals 0, which means that each a_k equals c_k (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.

Linearly independent — **Definition 2.15**

- A list $v_1, \dots, v_m$ of vectors in V is called *linearly independent* if the only choice of $a_1, \dots, a_m \in \mathbb{F}$ that makes

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

  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 $v_1, \dots, v_m$.

Example 2.16 **Linearly independent lists**

(a) To see that the list $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$ is linearly independent in $\mathbb{F}^4$, suppose $a_1, a_2, a_3 \in \mathbb{F}$ and

$$a_1(1, 0, 0, 0) + a_2(0, 1, 0, 0) + a_3(0, 0, 1, 0) = (0, 0, 0, 0).$$

Thus

$$(a_1, a_2, a_3, 0) = (0, 0, 0, 0).$$

Hence $a_1 = a_2 = a_3 = 0$. Thus the list $(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0)$ is linearly independent in $\mathbb{F}^4$.

(b) Suppose m is a nonnegative integer. To see that the list $1, z, \dots, z^m$ is linearly independent in $\mathbb{F}[X]$, suppose $a_0, a_1, \dots, a_m \in \mathbb{F}$ and

$$a_0 + a_1 z + \cdots + a_m z^m = 0,$$

where we think of both sides as elements of $\mathbb{F}[X]$. Then

$$a_0 + a_1 z + \cdots + a_m z^m = 0$$

for all $z \in \mathbb{F}$. As discussed earlier (and as follows from 4.8), this implies that $a_0 = a_1 = \cdots = a_m = 0$. Thus $1, z, \dots, z^m$ is a linearly independent list in $\mathbb{F}[X]$.

(c) A list of length one in a vector space is linearly independent if and only if the vector in the list is not 0.

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

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

Linearly dependent — Definition 2.17

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

Example 2.18 **Linearly dependent lists**

- $(2, 3, 1), (1, -1, 2), (7, 3, 8)$ is linearly dependent in $\mathbb{F}^3$ 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 $\mathbb{F}^3$ 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 next lemma is a terrific tool. It states that given a linearly dependent list of vectors, one of the vectors is in the span of the previous ones. Furthermore, we can throw out that vector without changing the span of the original list.

Linear dependence lemma — Result 2.19

Suppose $v_1, \ldots, v_m$ is a linearly dependent list in V. Then there exists $k \in \{1, 2, \ldots, m\}$ such that
$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$
Furthermore, if k satisfies the condition above and the kth term is removed from $v_1, \ldots, v_m$, then 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 $a_1, \ldots, a_m \in \mathbb{F}$, not all 0, such that
$$a_1 v_1 + \cdots + a_m v_m = 0.$$
Let k be the largest element of $\{1, \ldots, m\}$ such that $a_k \neq 0$. Then
$$v_k = -\frac{a_1}{a_k} v_1 - \cdots - \frac{a_{k-1}}{a_k} v_{k-1},$$
which proves that $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$, as desired.

Now suppose k is any element of $\{1, \dots, m\}$ such that $v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$. Let $b_1, \dots, b_{k-1} \in \mathbb{F}$ be such that

$$v_k = b_1 v_1 + \cdots + b_{k-1} v_{k-1}. \tag{2.20}$$

Suppose $u \in \operatorname{span}(v_1, \dots, v_m)$. Then there exist $c_1, \dots, c_m \in \mathbb{F}$ such that

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

In the equation above, we can replace v_k with the right side of 2.20, which shows that u is in the span of the list obtained by removing the kth term from $v_1, \dots, v_m$. Thus removing the kth term of the list $v_1, \dots, v_m$ does not change the span of the list. □

If $k = 1$ in the linear dependence lemma, then $v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$ means that $v_1 = 0$, because $\operatorname{span}() = \{0\}$. Note also that parts of the proof of the linear dependence lemma need to be modified if $k = 1$. In general, the proofs in the rest of the book will not call attention to special cases that must be considered involving lists of length 0, the subspace $\{0\}$, or other trivial cases for which the result is true but needs a slightly different proof. Be sure to check these special cases yourself.

Example 2.21 **Smallest k in linear dependence lemma**

Consider the list

$$(1, 2, 3), (6, 5, 4), (15, 16, 17), (8, 9, 7)$$

in $\mathbb{R}^3$. This list of length four is linearly dependent, as we will soon see. Thus the linear dependence lemma implies that there exists $k \in \{1, 2, 3, 4\}$ such that the kth vector in this list is a linear combination of the previous vectors in the list. Let's see how to find the smallest value of k that works.

Taking $k = 1$ in the linear dependence lemma works if and only if the first vector in the list equals 0. Because $(1, 2, 3)$ is not the 0 vector, we cannot take $k = 1$ for this list.

Taking $k = 2$ in the linear dependence lemma works if and only if the second vector in the list is a scalar multiple of the first vector. However, there does not exist $c \in \mathbb{R}$ such that $(6, 5, 4) = c(1, 2, 3)$. Thus we cannot take $k = 2$ for this list.

Taking $k = 3$ in the linear dependence lemma works if and only if the third vector in the list is a linear combination of the first two vectors. Thus for the list in this example, we want to know whether there exist $a, b \in \mathbb{R}$ such that

$$(15, 16, 17) = a(1, 2, 3) + b(6, 5, 4).$$

The equation above is equivalent to a system of three linear equations in the two unknowns a, b. Using Gaussian elimination or appropriate software, we find that $a = 3$, $b = 2$ is a solution of the equation above, as you can verify. Thus for the list in this example, taking $k = 3$ is the smallest value of k that works in the linear dependence lemma.

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

Length of linearly independent list $\leq$ length of spanning list Result 2.22
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 that $u_1, \dots, u_m$ 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 process described below with m steps; 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. Adjoining u_1 at the beginning of this list produces a linearly dependent list (because u_1 can be written as a linear combination of $w_1, \dots, w_n$). In other words, the list

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

is linearly dependent.

Thus by the linear dependence lemma (2.19), one of the vectors in the list above is a linear combination of the previous vectors in the list. We know that $u_1 \neq 0$ because the list $u_1, \dots, u_m$ is linearly independent. Thus u_1 is not in the span of the previous vectors in the list above (because u_1 is not in $\{0\}$, which is the span of the empty list). Hence the linear dependence lemma implies that 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 k, for $k = 2, \dots, m$

The list B (of length n) from step $k - 1$ spans V. In particular, u_k is in the span of the list B. Thus the list of length $(n + 1)$ obtained by adjoining u_k to B, placing it just after $u_1, \dots, u_{k-1}$, is linearly dependent. By the linear dependence lemma (2.19), one of the vectors in this list is in the span of the previous ones, and because $u_1, \dots, u_k$ is linearly independent, this vector cannot be one of the u's.

Hence there still must be at least one remaining w at this step. We can remove from our new list (after adjoining u_k in the proper place) a w that is a linear combination of the previous vectors in the list, so that the new list B (of length n) consisting of $u_1, \dots, u_k$ 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.

Example 2.23 **No list of length 4 is linearly independent in $\mathbb{R}^3$**

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

For example, we now know that $(1, 2, 3), (4, 5, 8), (9, 6, 7), (-3, 2, 8)$, which is a list of length four, is not linearly independent in $\mathbb{R}^3$.

Example 2.24 **No list of length 3 spans $\mathbb{R}^4$**

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

For example, we now know that $(1, 2, 3, -5), (4, 5, 8, 3), (9, 6, 7, -1)$, which is a list of length three, does not span $\mathbb{R}^4$.

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.

Finite-dimensional subspaces **Result 2.25**
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 multistep construction.

Step 1
If $U = \{0\}$, then U is finite-dimensional and we are done. If $U \neq \{0\}$, then choose a nonzero vector $u_1 \in U$.

Step k
If $U = \operatorname{span}(u_1, \ldots, u_{k-1})$, then U is finite-dimensional and we are done. If $U \neq \operatorname{span}(u_1, \ldots, u_{k-1})$, then choose a vector $u_k \in U$ such that

$$u_k \notin \operatorname{span}(u_1, \ldots, u_{k-1}).$$

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 (2.19). This linearly independent list cannot be longer than any spanning list of V (by 2.22). Thus the process eventually terminates, which means that U is finite-dimensional. □

Exercises

2A.1 Find a list of four distinct vectors in $\mathbb{F}^3$ whose span equals

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

2A.2 Prove or give a counterexample: If v_1, v_2, v_3, v_4 spans V, then the list

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

also spans V.

2A.3 Suppose $v_1, \ldots, v_m$ is a list of vectors in V. For $k \in \{1, \ldots, m\}$, let

$$w_k = v_1 + \cdots + v_k.$$

Show that $\operatorname{span}(v_1, \ldots, v_m) = \operatorname{span}(w_1, \ldots, w_m)$.

2A.4

(a) Show that a list of length one in a vector space is linearly independent if and only if the vector in the list is not 0.

(b) Show that a list of length two in a vector space is linearly independent if and only if neither of the two vectors in the list is a scalar multiple of the other.

2A.5 Find a number t such that

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

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

2A.6 Show that the list $(2, 3, 1), (1, -1, 2), (7, 3, c)$ is linearly dependent in $\mathbb{F}^3$ if and only if $c = 8$.

2A.7

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

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

2A.8 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.

2A.9 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.

2A.10 Prove or give a counterexample: If $v_1, v_2, \ldots, v_m$ is a linearly independent list of vectors in V and $\lambda \in \mathbb{F}$ with $\lambda \neq 0$, then $\lambda v_1, \lambda v_2, \ldots, \lambda v_m$ is linearly independent.

2A.11 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 the list $v_1 + w_1, \ldots, v_m + w_m$ is linearly independent.

2A.12 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 $w \in \operatorname{span}(v_1, \ldots, v_m)$.

2A.13 Suppose $v_1, \ldots, v_m$ is linearly independent in V and $w \in V$. Show that

$$v_1, \ldots, v_m, w \text{ is linearly independent} \iff w \notin \operatorname{span}(v_1, \ldots, v_m).$$

2A.14 Suppose $v_1, \ldots, v_m$ is a list of vectors in V. For $k \in \{1, \ldots, m\}$, let

$$w_k = v_1 + \cdots + v_k.$$

Show that the list $v_1, \ldots, v_m$ is linearly independent if and only if the list $w_1, \ldots, w_m$ is linearly independent.

2A.15 Explain why there does not exist a list of six polynomials that is linearly independent in $\mathbb{F}[X]_{\leq 4}$.

2A.16 Explain why no list of four polynomials spans $\mathbb{F}[X]_{\leq 4}$.

2A.17 Prove that V is infinite-dimensional if and only if there is a sequence $v_1, v_2, \ldots$ of vectors in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m.

2A.18 Prove that $\mathbb{F}^\infty$ is infinite-dimensional.

2A.19 Prove that the real vector space of all continuous real-valued functions on the interval $[0, 1]$ is infinite-dimensional.

2A.20 Suppose $p_0, p_1, \ldots, p_m$ are polynomials in $\mathbb{F}[X]_{\leq m}$ such that $p_k(2) = 0$ for each $k \in \{0, \ldots, m\}$. Prove that $p_0, p_1, \ldots, p_m$ is not linearly independent in $\mathbb{F}[X]_{\leq m}$.

2B Bases

In the previous section, we discussed linearly independent lists and we also discussed spanning lists. Now we bring these concepts together by considering lists that have both properties.

Basis — Definition 2.26

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

Example 2.27 **Bases**

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

(b) The list $(1, 2), (3, 5)$ is a basis of $\mathbb{F}^2$. Note that this list has length two, which is the same as the length of the standard basis of $\mathbb{F}^2$. In the next section, we will see that this is not a coincidence.

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

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

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

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

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

(g) The list $1, z, \ldots, z^m$ is a basis of $\mathbb{F}[X]_{\leq m}$, called the *standard basis* of $\mathbb{F}[X]_{\leq m}$.

In addition to the standard basis, $\mathbb{F}^n$ has many other bases. For example,

$$(7, 5), (-4, 9) \quad \text{and} \quad (1, 2), (3, 5)$$

are both bases of $\mathbb{F}^2$.

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

Criterion for basis — Result 2.28

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

$$v = a_1 v_1 + \cdots + a_n v_n, \tag{2.29}$$

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

Proof. First suppose that $v_1, \ldots, v_n$ is a basis of V. Let $v \in V$. Because $v_1, \ldots, v_n$ spans V, there exist $a_1, \ldots, a_n \in \mathbb{F}$ such that 2.29 holds. To show that the representation in 2.29 is unique, suppose $c_1, \ldots, c_n$ 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 2.29, we get

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

This implies that each $a_k - c_k$ equals 0 (because $v_1, \ldots, v_n$ is linearly independent). Hence $a_1 = c_1, \ldots, a_n = c_n$. 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 2.29. This implies that the list $v_1, \ldots, v_n$ spans V. To show that $v_1, \ldots, v_n$ is linearly independent, suppose $a_1, \ldots, a_n \in \mathbb{F}$ are such that

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

The uniqueness of the representation 2.29 (taking $v = 0$) now implies that $a_1 = \cdots = a_n = 0$. Thus $v_1, \ldots, v_n$ 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 $\mathbb{F}^2$, 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 $\mathbb{F}^2$.

Every spanning list contains a basis — Result 2.30

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

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

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

Step 1

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

Step k

If v_k is in $\operatorname{span}(v_1, \ldots, v_{k-1})$, then delete v_k from the list B. If v_k is not in $\operatorname{span}(v_1, \ldots, v_{k-1})$, then 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 (2.19). Hence B is a basis of V. □

We now come to an important corollary of the previous result.

Basis of finite-dimensional vector space Result 2.31

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 2.30, 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.

Every linearly independent list extends to a basis Result 2.32

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

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

$$u_1, \ldots, u_m, w_1, \ldots, w_n$$

spans V. Applying the procedure of the proof of 2.30 to reduce this list to a basis of V produces a basis consisting of the vectors $u_1, \ldots, u_m$ and some of the w's (none of the u's get deleted in this procedure because $u_1, \ldots, u_m$ is linearly independent). □

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

$$(2, 3, 4), (9, 6, 8), (0, 1, 0),$$

which is a basis of $\mathbb{F}^3$.

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.

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

Every subspace of V is part of a direct sum equal to V **Result 2.33**
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 2.25). Thus there is a basis $u_1, \dots, u_m$ of U (by 2.31). 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 (by 2.32). Let $W = \operatorname{span}(w_1, \dots, w_n)$.

To prove that $V = U \oplus W$, by 1.46 we only need to 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 $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$ such that

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

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 $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$ 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

2B.1 Find all vector spaces that have exactly one basis.

2B.2 Verify all assertions in Example 2.27.

2B.3

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

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

Find a basis of U.

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

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

2B.4
(a) Let U be the subspace of $\mathbb{C}^5$ defined by

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

Find a basis of U.

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

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

2B.5 Suppose V is finite-dimensional and U, W are subspaces of V such that $V = U + W$. Prove that there exists a basis of V consisting of vectors in $U \cup W$.

2B.6 Prove or give a counterexample: If p_0, p_1, p_2, p_3 is a list in $\mathbb{F}[X]_{\leq 3}$ such that none of the polynomials p_0, p_1, p_2, p_3 has degree 2, then p_0, p_1, p_2, p_3 is not a basis of $\mathbb{F}[X]_{\leq 3}$.

2B.7 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.

2B.8 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.

2B.9 Suppose $v_1, \ldots, v_m$ is a list of vectors in V. For $k \in \{1, \ldots, m\}$, let

$$w_k = v_1 + \cdots + v_k.$$

Show that $v_1, \ldots, v_m$ is a basis of V if and only if $w_1, \ldots, w_m$ is a basis of V.

2B.10 Suppose U and W are subspaces of V such that $V = U \oplus W$. Suppose also that $u_1, \ldots, u_m$ is a basis of U and $w_1, \ldots, w_n$ is a basis of W. Prove that

$$u_1, \ldots, u_m, w_1, \ldots, w_n$$

is a basis of V.

2B.11 Suppose V is a real vector space. Show that if $v_1, \ldots, v_n$ is a basis of V (as a real vector space), then $v_1, \ldots, v_n$ is also a basis of the complexification $V_{\mathbb{C}}$ (as a complex vector space).

See Exercise 8 in Section 1B for the definition of the complexification $V_{\mathbb{C}}$.

2C 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 $\mathbb{F}^n$ to equal n. Notice that the standard basis

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

of $\mathbb{F}^n$ 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.

Basis length does not depend on basis — Result 2.34

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 2.22). 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.

Dimension, $\dim V$ — Definition 2.35

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

Example 2.36 Dimensions

- $\dim \mathbb{F}^n = n$ because the standard basis of $\mathbb{F}^n$ has length n.
- $\dim \mathbb{F}[X]_{\le m} = m + 1$ because the standard basis $1, z, \ldots, z^m$ of $\mathbb{F}[X]_{\le m}$ has length $m + 1$.
- If $U = \{(x, x, y) \in \mathbb{F}^3 : x, y \in \mathbb{F}\}$, then $\dim U = 2$ because $(1, 1, 0), (0, 0, 1)$ is a basis of U.
- If $U = \{(x, y, z) \in \mathbb{F}^3 : x + y + z = 0\}$, then $\dim U = 2$ because the list $(1, -1, 0), (1, 0, -1)$ is a basis of U.

Every subspace of a finite-dimensional vector space is finite-dimensional (by 2.25) and so has a dimension. The next result gives the expected inequality about the dimension of a subspace.

Dimension of a subspace — Result 2.37

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 2.22 to conclude that $\dim U \leq \dim V$. □

The real vector space $\mathbb{R}^2$ has dimension two; the complex vector space $\mathbb{C}$ has dimension one. As sets, $\mathbb{R}^2$ can be identified with $\mathbb{C}$ (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 $\mathbb{F}$ cannot be neglected.

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 only need to check that it satisfies one of the two required properties. First we prove that every linearly independent list of the right length is a basis.

Linearly independent list of the right length is a basis — Result 2.38

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

Proof. Suppose $\dim V = n$ and $v_1, \ldots, v_n$ is linearly independent in V. The list $v_1, \ldots, v_n$ can be extended to a basis of V (by 2.32). 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 $v_1, \ldots, v_n$. Thus $v_1, \ldots, v_n$ is a basis of V, as desired. □

The next result is a useful consequence of the previous result.

Subspace of full dimension equals the whole space — Result 2.39

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

Proof. Let $u_1, \ldots, u_n$ be a basis of U. Thus $n = \dim U$, and by hypothesis we also have $n = \dim V$. Thus $u_1, \ldots, u_n$ is a linearly independent list of vectors in V (because it is a basis of U) of length $\dim V$. From 2.38, we see that $u_1, \ldots, u_n$ is a basis of V. In particular every vector in V is a linear combination of $u_1, \ldots, u_n$. Thus $U = V$. □

Example 2.40 **A basis of $\mathbb{F}^2$**

Consider the list $(5, 7), (4, 3)$ of vectors in $\mathbb{F}^2$. This list of length two is linearly independent in $\mathbb{F}^2$ (because neither vector is a scalar multiple of the other). Note that $\mathbb{F}^2$ has dimension two. Thus 2.38 implies that the linearly independent list $(5, 7), (4, 3)$ of length two is a basis of $\mathbb{F}^2$ (we do not need to bother checking

Example 2.40 *continued*

that it spans $\mathbb{F}^2$).

Example 2.41 **A basis of a subspace of $\mathbb{R}[X]_{\le 3}$**

Let U be the subspace of $\mathbb{R}[X]_{\le 3}$ defined by

$$U = \{p \in \mathbb{R}[X]_{\le 3} : p'(5) = 0\}.$$

To find a basis of U, first note that each of the polynomials 1, $(x-5)^2$, and $(x-5)^3$ is in U.

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

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

for every $x \in \mathbb{R}$. 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 $3 \le \dim U$. Hence

$$3 \le \dim U \le \dim \mathbb{R}[X]_{\le 3} = 4,$$

where we have used 2.37.

The polynomial x is not in U because its derivative is the constant function 1. Thus $U \ne \mathbb{R}[X]_{\le 3}$. Hence $\dim U \ne 4$ (by 2.39). The inequality above now implies that $\dim U = 3$. Thus the linearly independent list $1, (x-5)^2, (x-5)^3$ in U has length $\dim U$ and hence is a basis of U (by 2.38).

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

Spanning list of the right length is a basis Result 2.42

Suppose V is finite-dimensional. Then every list of vectors in V that spans V and has 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 2.30). 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 $v_1, \dots, v_n$. Thus $v_1, \dots, v_n$ 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.

Dimension of a sum **Result 2.43**

If V_1 and V_2 are subspaces of a finite-dimensional vector space, then

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

Proof. Let $v_1, \dots, v_m$ be a basis of $V_1 \cap V_2$; thus $\dim(V_1 \cap V_2) = m$. Because $v_1, \dots, v_m$ is a basis of $V_1 \cap V_2$, it is linearly independent in V_1. Hence this list can be extended to a basis $v_1, \dots, v_m, u_1, \dots, u_j$ of V_1 (by 2.32). Thus $\dim V_1 = m + j$. Also extend $v_1, \dots, v_m$ to a basis $v_1, \dots, v_m, w_1, \dots, w_k$ of V_2; thus $\dim V_2 = m + k$.

We will show that

$$v_1, \dots, v_m, u_1, \dots, u_j, w_1, \dots, w_k \tag{2.44}$$

is a basis of $V_1 + V_2$. This will complete the proof, because then we will have

$$\begin{aligned}\dim(V_1 + V_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).\end{aligned}$$

The list 2.44 is contained in $V_1 \cup V_2$ and thus is contained in $V_1 + V_2$. The span of this list contains V_1 and contains V_2 and hence is equal to $V_1 + V_2$. Thus to show that 2.44 is a basis of $V_1 + V_2$ we only need to show that it is linearly independent.

To prove that 2.44 is linearly independent, suppose

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j + c_1 w_1 + \cdots + c_k w_k = 0,$$

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

$$c_1 w_1 + \cdots + c_k w_k = -a_1 v_1 - \cdots - a_m v_m - b_1 u_1 - \cdots - b_j u_j, \tag{2.45}$$

which shows that $c_1 w_1 + \cdots + c_k w_k \in V_1$. All the w's are in V_2, so this implies that $c_1 w_1 + \cdots + c_k w_k \in V_1 \cap V_2$. Because $v_1, \dots, v_m$ is a basis of $V_1 \cap V_2$, we have

$$c_1 w_1 + \cdots + c_k w_k = d_1 v_1 + \cdots + d_m v_m$$

for some scalars $d_1, \dots, d_m$. But $v_1, \dots, v_m, w_1, \dots, w_k$ is linearly independent, so the last equation implies that all the c's (and d's) equal 0. Thus 2.45 becomes the equation

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j = 0.$$

Because the list $v_1, \dots, v_m, u_1, \dots, u_j$ is linearly independent, this equation implies that all the a's and b's are 0, completing the proof. □

For S a finite set, let $\#S$ denote the number of elements of S. The table below compares finite sets with finite-dimensional vector spaces, showing the analogy between $\#S$ (for sets) and $\dim V$ (for vector spaces), as well as the analogy between unions of subsets (in the context of sets) and sums of subspaces (in the context of vector spaces).

sets	vector spaces
S is a finite set	V is a finite-dimensional vector space
$\#S$	$\dim V$
for subsets S_1, S_2 of S, the union $S_1 \cup S_2$ is the smallest subset of S containing S_1 and S_2	for subspaces V_1, V_2 of V, the sum $V_1 + V_2$ is the smallest subspace of V containing V_1 and V_2
$\#(S_1 \cup S_2)$ $= \#S_1 + \#S_2 - \#(S_1 \cap S_2)$	$\dim(V_1 + V_2)$ $= \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2)$
$\#(S_1 \cup S_2) = \#S_1 + \#S_2$ $\iff S_1 \cap S_2 = \varnothing$	$\dim(V_1 + V_2) = \dim V_1 + \dim V_2$ $\iff V_1 \cap V_2 = \{0\}$
$S_1 \cup \cdots \cup S_m$ is a disjoint union $\iff$ $\#(S_1 \cup \cdots \cup S_m) = \#S_1 + \cdots + \#S_m$	$V_1 + \cdots + V_m$ is a direct sum $\iff$ $\dim(V_1 + \cdots + V_m)$ $= \dim V_1 + \cdots + \dim V_m$

The last row above focuses on the analogy between disjoint unions (for sets) and direct sums (for vector spaces). The proof of the result in the last box above will be given in 3.94.

You should be able to find results about sets that correspond, via analogy, to the results about vector spaces in Exercises 12 through 18.

Exercises

2C.1 Show that the subspaces of $\mathbb{R}^2$ are precisely $\{0\}$, all lines in $\mathbb{R}^2$ containing the origin, and $\mathbb{R}^2$.

2C.2 Show that the subspaces of $\mathbb{R}^3$ are precisely $\{0\}$, all lines in $\mathbb{R}^3$ containing the origin, all planes in $\mathbb{R}^3$ containing the origin, and $\mathbb{R}^3$.

2C.3

(a) Let $U = \{p \in \mathbb{F}[X]_{\leq 4} : p(6) = 0\}$. Find a basis of U.

(b) Extend the basis in (a) to a basis of $\mathbb{F}[X]_{\leq 4}$.

(c) Find a subspace W of $\mathbb{F}[X]_{\leq 4}$ such that $\mathbb{F}[X]_{\leq 4} = U \oplus W$.

2C.4

(a) Let $U = \{p \in \mathbb{R}[X]_{\leq 4} : p''(6) = 0\}$. Find a basis of U.

(b) Extend the basis in (a) to a basis of $\mathbb{R}[X]_{\leq 4}$.

(c) Find a subspace W of $\mathbb{R}[X]_{\leq 4}$ such that $\mathbb{R}[X]_{\leq 4} = U \oplus W$.

2C.5

(a) Let $U = \{p \in \mathbb{F}[X]_{\leq 4} : p(2) = p(5)\}$. Find a basis of U.

(b) Extend the basis in (a) to a basis of $\mathbb{F}[X]_{\leq 4}$.

(c) Find a subspace W of $\mathbb{F}[X]_{\leq 4}$ such that $\mathbb{F}[X]_{\leq 4} = U \oplus W$.

2C.6
(a) Let $U = \{p \in \mathbb{F}[X]_{\le 4} : p(2) = p(5) = p(6)\}$. Find a basis of U.

(b) Extend the basis in (a) to a basis of $\mathbb{F}[X]_{\le 4}$.

(c) Find a subspace W of $\mathbb{F}[X]_{\le 4}$ such that $\mathbb{F}[X]_{\le 4} = U \oplus W$.

2C.7
(a) Let $U = \left\{p \in \mathbb{R}[X]_{\le 4} : \int_{-1}^{1} p = 0\right\}$. Find a basis of U.

(b) Extend the basis in (a) to a basis of $\mathbb{R}[X]_{\le 4}$.

(c) Find a subspace W of $\mathbb{R}[X]_{\le 4}$ such that $\mathbb{R}[X]_{\le 4} = U \oplus W$.

2C.8 Suppose $v_1, \dots, v_m$ is linearly independent in V and $w \in V$. Prove that

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

2C.9 Suppose m is a positive integer and $p_0, p_1, \dots, p_m \in \mathbb{F}[X]$ are such that each p_k has degree k. Prove that $p_0, p_1, \dots, p_m$ is a basis of $\mathbb{F}[X]_{\le m}$.

2C.10 Suppose m is a positive integer. For $0 \le k \le m$, let

$$p_k(x) = x^k(1 - x)^{m-k}.$$

Show that $p_0, \dots, p_m$ is a basis of $\mathbb{F}[X]_{\le m}$.

The basis in this exercise leads to what are called ***Bernstein polynomials****. You can do a web search to learn how Bernstein polynomials are used to approximate continuous functions on* $[0, 1]$.

2C.11 Suppose U and W are both four-dimensional subspaces of $\mathbb{C}^6$. Prove that there exist two vectors in $U \cap W$ such that neither of these vectors is a scalar multiple of the other.

2C.12 Suppose that U and W are subspaces of $\mathbb{R}^8$ such that $\dim U = 3$, $\dim W = 5$, and $U + W = \mathbb{R}^8$. Prove that $\mathbb{R}^8 = U \oplus W$.

2C.13 Suppose U and W are both five-dimensional subspaces of $\mathbb{R}^9$. Prove that $U \cap W \ne \{0\}$.

2C.14 Suppose V is a ten-dimensional vector space and V_1, V_2, V_3 are subspaces of V with $\dim V_1 = \dim V_2 = \dim V_3 = 7$. Prove that $V_1 \cap V_2 \cap V_3 \ne \{0\}$.

2C.15 Suppose V is finite-dimensional and V_1, V_2, V_3 are subspaces of V with $\dim V_1 + \dim V_2 + \dim V_3 > 2 \dim V$. Prove that $V_1 \cap V_2 \cap V_3 \ne \{0\}$.

2C.16 Suppose V is finite-dimensional and U is a subspace of V with $U \neq V$. Let $n = \dim V$ and $m = \dim U$. Prove that there exist $n - m$ subspaces of V, each of dimension $n - 1$, whose intersection equals U.

2C.17 Suppose that $V_1, \dots, V_m$ are finite-dimensional subspaces of V. Prove that $V_1 + \cdots + V_m$ is finite-dimensional and

$$\dim(V_1 + \cdots + V_m) \leq \dim V_1 + \cdots + \dim V_m.$$

The inequality above is an equality if and only if $V_1 + \cdots + V_m$ is a direct sum, as will be shown in 3.94.

2C.18 Suppose V is finite-dimensional, with $\dim V = n \geq 1$. Prove that there exist one-dimensional subspaces $V_1, \dots, V_n$ of V such that

$$V = V_1 \oplus \cdots \oplus V_n.$$

2C.19 Explain why you might guess, motivated by analogy with the formula for the number of elements in the union of three finite sets, that if V_1, V_2, V_3 are subspaces of a finite-dimensional vector space, then

$$\begin{aligned}&\dim(V_1 + V_2 + V_3)\\ &\quad = \dim V_1 + \dim V_2 + \dim V_3\\ &\qquad - \dim(V_1 \cap V_2) - \dim(V_1 \cap V_3) - \dim(V_2 \cap V_3)\\ &\qquad + \dim(V_1 \cap V_2 \cap V_3).\end{aligned}$$

Then either prove the formula above or give a counterexample.

2C.20 Prove that if V_1, V_2, and V_3 are subspaces of a finite-dimensional vector space, then

$$\begin{aligned}&\dim(V_1 + V_2 + V_3)\\ &\quad = \dim V_1 + \dim V_2 + \dim V_3\\ &\qquad - \frac{\dim(V_1 \cap V_2) + \dim(V_1 \cap V_3) + \dim(V_2 \cap V_3)}{3}\\ &\qquad - \frac{\dim\big((V_1 + V_2) \cap V_3\big) + \dim\big((V_1 + V_3) \cap V_2\big) + \dim\big((V_2 + V_3) \cap V_1\big)}{3}.\end{aligned}$$

The formula above may seem strange because the right side does not look like an integer.

I at once gave up my former occupations, set down natural history and all its progeny as a deformed and abortive creation, and entertained the greatest disdain for a would-be science which could never even step within the threshold of real knowledge. In this mood I betook myself to the mathematics and the branches of study appertaining to that science as being built upon secure foundations, and so worthy of my consideration.

—*Frankenstein*, Mary Wollstonecraft Shelley

3 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.

We will frequently use the powerful fundamental theorem of linear maps, which states that the dimension of the domain of a linear map equals the dimension of the subspace that gets sent to 0 plus the dimension of the range. This will imply the striking result that a linear map from a finite-dimensional vector space to itself is one-to-one if and only if its range is the whole space.

A major concept that we will introduce in this chapter is the matrix associated with a linear map and with a basis of the domain space and a basis of the target space. This correspondence between linear maps and matrices provides much insight into key aspects of linear algebra.

This chapter concludes by introducing product, quotient, and dual spaces.

In this chapter we will need additional vector spaces, which we call U and W, in addition to V. Thus our standing assumptions are now as follows.

Standing assumptions for this chapter — Definition

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

3A Vector Space of Linear Maps

Definition and Examples of Linear Maps

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

Linear map — Definition 3.1

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 $\lambda \in \mathbb{F}$ and all $v \in V$.

Definition 3.1 *continued*

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

Note that for linear maps we often use the notation Tv as well as the usual function notation $T(v)$.

Notation 3.2 $\mathrm{Hom}(V, W)$, $\mathrm{End}(V)$

- The set of linear maps from V to W is denoted by $\mathrm{Hom}(V, W)$.
- The set of linear maps from V to V is denoted by $\mathrm{End}(V)$. In other words, $\mathrm{End}(V) = \mathrm{Hom}(V, V)$.

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

Example 3.3 **Linear maps**

zero
In addition to its other uses, we let the symbol 0 denote the linear map that takes every element of some vector space to the additive identity of another (or possibly the same) vector space. To be specific, $0 \in \mathrm{Hom}(V, W)$ 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 operator
The *identity operator*, denoted by I, is the linear map on some vector space that takes each element to itself. To be specific, $I \in \mathrm{End}(V)$ is defined by

$$Iv = v.$$

differentiation
Define $D \in \mathrm{End}(\mathbb{R}[X])$ 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 $T \in \mathrm{Hom}(\mathbb{R}[X], \mathbb{R})$ by

$$Tp = \int_0^1 p.$$

Example 3.3
continued

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 a linear map $T \in \text{End}(\mathbb{R}[X])$ by

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

for each $x \in \mathbb{R}$.

backward shift
Recall that $\mathbb{F}^\infty$ denotes the vector space of all sequences of elements of $\mathbb{F}$. Define a linear map $T \in \text{End}(\mathbb{F}^\infty)$ by

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

from $\mathbb{R}^3$ to $\mathbb{R}^2$
Define a linear map $T \in \text{Hom}(\mathbb{R}^3, \mathbb{R}^2)$ by

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

from $\mathbb{F}^n$ to $\mathbb{F}^m$
To generalize the previous example, let m and n be positive integers, let $A_{j,k} \in \mathbb{F}$ for each $j = 1, \dots, m$ and each $k = 1, \dots, n$, and define a linear map $T \in \text{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ 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 $\mathbb{F}^n$ to $\mathbb{F}^m$ is of this form.

composition
Fix a polynomial $q \in \mathbb{R}[X]$. Define a linear map $T \in \text{End}(\mathbb{R}[X])$ by

$$(Tp)(x) = p\big(q(x)\big).$$

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.

Result 3.4

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

$$Tv_k = w_k$$

for each $k = 1, \dots, n$.

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 $\mathbb{F}$. The list $v_1, \ldots, v_n$ is a basis of V. 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 k, taking $c_k = 1$ and the other c's equal to 0 in the equation above shows that $T v_k = w_k$.

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 $\lambda \in \mathbb{F}$ 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 $T \in \operatorname{Hom}(V, W)$ and that $T v_k = w_k$ for each $k = 1, \ldots, n$. Let $c_1, \ldots, c_n \in \mathbb{F}$. Then the homogeneity of T implies that $T(c_k v_k) = c_k w_k$ for each $k = 1, \ldots, n$. 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, as desired. □

Linear maps are pervasive throughout mathematics. However, they are not as ubiquitous as imagined by people who seem to think cos is a linear map from $\mathbb{R}$ to $\mathbb{R}$ when they incorrectly write that $\cos(x + y)$ equals $\cos x + \cos y$ and that $\cos 2x$ equals $2 \cos x$.

Algebraic Operations on $\operatorname{Hom}(V, W)$

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

Addition and scalar multiplication on $\operatorname{Hom}(V, W)$ **Definition 3.5**

Suppose $S, T \in \operatorname{Hom}(V, W)$ and $\lambda \in \mathbb{F}$. 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 $S, T \in \text{Hom}(V, W)$ and $\lambda \in \mathbb{F}$, then $S + T \in \text{Hom}(V, W)$ and $\lambda T \in \text{Hom}(V, W)$.

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

$\text{Hom}(V, W)$ is a vector space — Result 3.6

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

The routine proof of the result above is left to the reader. Note that the additive identity of $\text{Hom}(V, W)$ is the zero linear map defined in Example 3.3.

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, as in the next definition.

Product of linear maps — Definition 3.7

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

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

for all $u \in U$.

Thus ST is just the usual composition $S \circ T$ of two functions, but when both functions are linear, we usually write ST instead of $S \circ T$. The product notation ST helps make the distributive properties (see next result) seem natural.

Note that ST is defined only when T maps into the domain of S. You should verify that ST is indeed a linear map from U to W whenever $T \in \text{Hom}(U, V)$ and $S \in \text{Hom}(V, W)$.

Algebraic properties of products of linear maps — Result 3.8

associativity
$(T_1T_2)T_3 = T_1(T_2T_3)$ whenever T_1, T_2, and T_3 are linear maps such that the products make sense (meaning T_3 maps into the domain of T_2, and T_2 maps into the domain of T_1).

identity
$TI = IT = T$ whenever $T \in \text{Hom}(V, W)$; here the first I is the identity operator on V, and the second I is the identity operator on W.

distributive properties
$(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$ whenever $T, T_1, T_2 \in \text{Hom}(U, V)$ and $S, S_1, S_2 \in \text{Hom}(V, W)$.

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.

Example 3.9 **Two noncommuting linear maps from $\mathbb{R}[X]$ to $\mathbb{R}[X]$**

Suppose $D \in \text{End}(\mathbb{R}[X])$ is the differentiation map defined in Example 3.3 and $T \in \text{End}(\mathbb{R}[X])$ is the multiplication by x^2 map defined earlier in this section. Then

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

Thus $TD \neq DT$—differentiating and then multiplying by x^2 is not the same as multiplying by x^2 and then differentiating.

Linear maps take 0 to 0 **Result 3.10**

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 $T(0) = 0$. □

Suppose $m, b \in \mathbb{R}$. The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = mx + b$$

is a linear map if and only if $b = 0$ (use 3.10). Thus the linear functions of high school algebra are not the same as linear maps in the context of linear algebra.

Exercises

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

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

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

3A.2 Suppose $b, c \in \mathbb{R}$. Define $T : \mathbb{R}[X] \to \mathbb{R}^2$ by

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

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

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

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

for every $(x_1, \dots, x_n) \in \mathbb{F}^n$.

This exercise shows that the linear map T has the form promised in the second to last item of Example 3.3.

3A.4 Suppose $T \in \text{Hom}(V, W)$ 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.

3A.5 Prove that $\text{Hom}(V, W)$ is a vector space, as was asserted in 3.6.

3A.6 Prove that multiplication of linear maps has the associative, identity, and distributive properties asserted in 3.8.

3A.7 Show that every linear map from a one-dimensional vector space to itself is multiplication by some scalar. More precisely, prove that if $\dim V = 1$ and $T \in \text{End}(V)$, then there exists $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$ for all $v \in V$.

3A.8 Give an example of a function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ such that

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

for all $a \in \mathbb{R}$ and all $v \in \mathbb{R}^2$ but φ is not linear.

This exercise and the next exercise show that neither homogeneity nor additivity alone is enough to imply that a function is a linear map.

3A.9 Give an example of a function $\varphi\colon \mathbb{C} \to \mathbb{C}$ such that

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

for all $w, z \in \mathbb{C}$ but φ is not linear. (Here $\mathbb{C}$ is thought of as a complex vector space.)

There also exists a function $\varphi\colon \mathbb{R} \to \mathbb{R}$ 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.

3A.10 Prove or give a counterexample: If $q \in \mathbb{R}[X]$ and $T\colon \mathbb{R}[X] \to \mathbb{R}[X]$ is defined by $Tp = q \circ p$, then T is a linear map.

The function T defined here differs from the function T defined in the last bullet point of 3.3 by the order of the functions in the compositions.

3A.11 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that T is a scalar multiple of the identity if and only if $ST = TS$ for every $S \in \text{End}(V)$.

3A.12 Suppose U is a subspace of V with $U \neq V$. Suppose $S \in \text{Hom}(U, W)$ and $S \neq 0$ (which means that $Su \neq 0$ for some $u \in U$). Define $T: V \to W$ 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.

3A.13 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 $S \in \text{Hom}(U, W)$, then there exists $T \in \text{Hom}(V, W)$ such that $Tu = Su$ for all $u \in U$.

The result in this exercise is used in the proof of 3.125.

3A.14 Suppose V is finite-dimensional with $\dim V > 0$, and suppose W is infinite-dimensional. Prove that $\text{Hom}(V, W)$ is infinite-dimensional.

3A.15 Suppose $v_1, \ldots, v_m$ is a linearly dependent list of vectors in V. Suppose also that $W \neq \{0\}$. Prove that there exist $w_1, \ldots, w_m \in W$ such that no $T \in \text{Hom}(V, W)$ satisfies $Tv_k = w_k$ for each $k = 1, \ldots, m$.

3A.16 Suppose V is finite-dimensional with $\dim V > 1$. Prove that there exist $S, T \in \text{End}(V)$ such that $ST \neq TS$.

3A.17 Suppose V is finite-dimensional. Show that the only two-sided ideals of $\text{End}(V)$ are $\{0\}$ and $\text{End}(V)$.

*A subspace $\mathscr{E}$ of $\text{End}(V)$ is called a **two-sided ideal** of $\text{End}(V)$ if $TE \in \mathscr{E}$ and $ET \in \mathscr{E}$ for all $E \in \mathscr{E}$ and all $T \in \text{End}(V)$.*

3B 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.

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

Null space, $\ker T$ — **Definition 3.11**

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

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

Example 3.12 **Null space**

- If T is the zero map from V to W, meaning that $Tv = 0$ for every $v \in V$, then $\ker T = V$.
- Suppose $\varphi \in \text{Hom}(\mathbb{C}^3, \mathbb{C})$ is defined by $\varphi(z_1, z_2, z_3) = z_1 + 2z_2 + 3z_3$. Then $\ker \varphi$ equals $\{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 + 2z_2 + 3z_3 = 0\}$, which is a subspace of the domain of φ. We will soon see that the null space of each linear map is a subspace of its domain.
- Suppose $D \in \text{End}(\mathbb{R}[X])$ 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 that $T \in \text{End}(\mathbb{R}[X])$ 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 $x \in \mathbb{R}$ is the 0 polynomial. Thus $\ker T = \{0\}$.
- Suppose $T \in \text{End}(\mathbb{F}^\infty)$ is the backward shift defined by

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

  Then $T(x_1, x_2, x_3, \dots)$ equals 0 if and only if the numbers $x_2, x_3, \dots$ are all 0. Thus $\ker T = \{(a, 0, 0, \dots) : a \in \mathbb{F}\}$.

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.

The null space is a subspace **Result 3.13**

Suppose $T \in \text{Hom}(V, W)$. Then $\ker T$ is a subspace of V.

Proof. Because T is a linear map, $T(0) = 0$ (by 3.10). Thus $0 \in \ker T$.

Suppose $u, v \in \ker T$. Then

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

Hence $u + v \in \ker T$. Thus $\ker T$ is closed under addition.

Suppose $u \in \ker T$ and $\lambda \in \mathbb{F}$. Then

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

Hence $\lambda u \in \ker T$. Thus $\ker T$ is closed under scalar multiplication.

We have shown that $\ker T$ contains 0 and is closed under addition and scalar multiplication. Thus $\ker T$ 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.

Injective — **Definition 3.14**

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

We could rephrase the definition above to say that T is injective if $u \neq v$ implies that $Tu \neq Tv$. Thus T is injective if and only if it maps distinct inputs to distinct outputs.

The term *one-to-one* means the same as injective.

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 3.12, only multiplication by x^2 is injective (except that the zero map is injective in the special case $V = \{0\}$).

Injectivity $\iff$ null space equals $\{0\}$ — **Result 3.15**

Let $T \in \text{Hom}(V, W)$. Then T is injective if and only if $\ker T = \{0\}$.

Proof. First suppose T is injective. We want to prove that $\ker T = \{0\}$. We already know that $\{0\} \subseteq \ker T$ (by 3.10). To prove the inclusion in the other direction, suppose $v \in \ker T$. Then

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

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

To prove the implication in the other direction, now suppose $\ker T = \{0\}$. 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 $\ker T$, 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 linear map.

Range — **Definition 3.16**

For $T \in \text{Hom}(V, W)$, the *range* of T is the subset of W consisting of those vectors that are equal to Tv for some $v \in V$:

$$\text{im } T = \{Tv : v \in V\}.$$

Example 3.17 **Range**

Example 3.17 *continued*

- If T is the zero map from V to W, meaning that $Tv = 0$ for every $v \in V$, then $\operatorname{im} T = \{0\}$.
- Suppose $T \in \operatorname{Hom}(\mathbb{R}^2, \mathbb{R}^3)$ is defined by $T(x, y) = (2x, 5y, x + y)$. Then
$$\operatorname{im} T = \{(2x, 5y, x + y) : x, y \in \mathbb{R}\}.$$
Note that $\operatorname{im} T$ is a subspace of $\mathbb{R}^3$. We will soon see that the range of each element of $\operatorname{Hom}(V, W)$ is a subspace of W.
- Suppose $D \in \operatorname{End}(\mathbb{R}[X])$ is the differentiation map defined by $Dp = p'$. Because for every polynomial $q \in \mathbb{R}[X]$ there exists a polynomial $p \in \mathbb{R}[X]$ such that $p' = q$, the range of D is $\mathbb{R}[X]$.

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

The range is a subspace **Result 3.18**
If $T \in \operatorname{Hom}(V, W)$, then $\operatorname{im} T$ is a subspace of W.

Proof. Suppose $T \in \operatorname{Hom}(V, W)$. Then $T(0) = 0$ (by 3.10), which implies that $0 \in \operatorname{im} T$.

If $w_1, w_2 \in \operatorname{im} T$, 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 $w_1 + w_2 \in \operatorname{im} T$. Thus $\operatorname{im} T$ is closed under addition.

If $w \in \operatorname{im} T$ and $\lambda \in \mathbb{F}$, then there exists $v \in V$ such that $Tv = w$. Thus
$$T(\lambda v) = \lambda Tv = \lambda w.$$
Hence $\lambda w \in \operatorname{im} T$. Thus $\operatorname{im} T$ is closed under scalar multiplication.

We have shown that $\operatorname{im} T$ contains 0 and is closed under addition and scalar multiplication. Thus $\operatorname{im} T$ is a subspace of W (by 1.34). □

Surjective **Definition 3.19**
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 3.17, only the differentiation map is surjective (except that the zero map is surjective in the special case $W = \{0\}$).

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

Some people use the term *onto*, which means the same as surjective.

Example 3.20 **Surjectivity depends on the target space**

The differentiation map $D \in \mathrm{End}(\mathbb{R}[X]_{\le 5})$ defined by $Dp = p'$ is not surjective, because the polynomial x^5 is not in the range of D. However, the differentiation map $S \in \mathrm{Hom}(\mathbb{R}[X]_{\le 5}, \mathbb{R}[X]_{\le 4})$ defined by $Sp = p'$ is surjective, because its range equals $\mathbb{R}[X]_{\le 4}$, which is the vector space into which S maps.

Fundamental Theorem of Linear Maps

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

Fundamental theorem of linear maps **Result 3.21**

Suppose V is finite-dimensional and $T \in \mathrm{Hom}(V, W)$. Then $\operatorname{im} T$ is finite-dimensional and

$$\dim V = \dim \ker T + \dim \operatorname{im} T.$$

Proof. Let $u_1, \dots, u_m$ be a basis of $\ker T$; thus $\dim \ker T = m$. The linearly independent list $u_1, \dots, u_m$ can be extended to a basis

$$u_1, \dots, u_m, v_1, \dots, v_n$$

of V (by 2.32). Thus $\dim V = m + n$. To complete the proof, we need to show that $\operatorname{im} T$ is finite-dimensional and $\dim \operatorname{im} T = n$. We will do this by proving that $Tv_1, \dots, Tv_n$ is a basis of $\operatorname{im} T$.

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

$$v = a_1 u_1 + \cdots + a_m u_m + b_1 v_1 + \cdots + b_n v_n,$$

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

$$Tv = b_1 Tv_1 + \cdots + b_n Tv_n,$$

where the terms of the form Tu_k disappeared because each u_k is in $\ker T$. The last equation implies that the list $Tv_1, \dots, Tv_n$ spans $\operatorname{im} T$. In particular, $\operatorname{im} T$ is finite-dimensional.

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

$$c_1 Tv_1 + \cdots + c_n Tv_n = 0.$$

Then

$$T(c_1 v_1 + \cdots + c_n v_n) = 0.$$

Hence

$$c_1 v_1 + \cdots + c_n v_n \in \ker T.$$

Because $u_1, \dots, u_m$ spans $\ker T$, we can write

$$c_1 v_1 + \cdots + c_n v_n = d_1 u_1 + \cdots + d_m u_m,$$

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

Linear map to a lower-dimensional space is not injective — Result 3.22

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 $T \in \mathrm{Hom}(V, W)$. Then

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

where the first line above comes from the fundamental theorem of linear maps (3.21) and the second line follows from 2.37. The inequality above states that $\dim \ker T > 0$. This means that $\ker T$ contains vectors other than 0. Thus T is not injective (by 3.15). □

Example 3.23 **Linear map from $\mathbb{F}^4$ to $\mathbb{F}^3$ is not injective**

Define a linear map $T\colon \mathbb{F}^4 \to \mathbb{F}^3$ by

$$T(z_1, z_2, z_3, z_4) = \big(\sqrt{7}z_1 + \pi z_2 + z_4,\ 97z_1 + 3z_2 + 2z_3,\ z_2 + 6z_3 + 7z_4\big).$$

Because $\dim \mathbb{F}^4 > \dim \mathbb{F}^3$, we can use 3.22 to assert that T is not injective, without doing any calculations.

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.

Linear map to a higher-dimensional space is not surjective — Result 3.24

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 $T \in \mathrm{Hom}(V, W)$. Then

$$\begin{aligned} \dim \operatorname{im} T &= \dim V - \dim \ker T \\ &\leq \dim V \\ &< \dim W, \end{aligned}$$

where the equality above comes from the fundamental theorem of linear maps (3.21). The inequality above states that $\dim \operatorname{im} T < \dim W$. This means that $\operatorname{im} T$ cannot equal W. Thus T is not surjective. □

As we will soon see, 3.22 and 3.24 have important consequences in the theory of linear equations. The idea is to express questions about systems of linear equations in terms of linear maps. Let's begin by rephrasing in terms of linear maps the question of whether a homogeneous system of linear equations has a nonzero solution.

Fix positive integers m and n, and let $A_{j,k} \in \mathbb{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$. Consider the homogeneous system of linear equations

Homogeneous, in this context, means that the constant term on the right side of each equation below is 0.

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= 0 \\ &\vdots \\ \sum_{k=1}^{n} A_{m,k} x_k &= 0. \end{aligned}$$

Clearly $x_1 = \cdots = x_n = 0$ is a solution of the system of equations above; the question here is whether any other solutions exist.

Define $T: \mathbb{F}^n \to \mathbb{F}^m$ by

$$T(x_1, \dots, x_n) = \Big(\sum_{k=1}^{n} A_{1,k} x_k, \dots, \sum_{k=1}^{n} A_{m,k} x_k \Big). \tag{3.25}$$

The equation $T(x_1, \dots, x_n) = 0$ (the 0 here is the additive identity in $\mathbb{F}^m$, 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 $\ker T$ is strictly bigger than $\{0\}$, which is equivalent to T not being injective (by 3.15). The next result gives an important condition for ensuring that T is not injective.

Result 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 discussion above. Thus T is a linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$, and we have a homogeneous system of m linear equations with n variables $x_1, \dots, x_n$. From 3.22 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.

Now we consider the question of whether a system of linear equations has no solutions for some choice of the constant terms. To rephrase this question in terms of a linear map, fix positive integers m and n, and let $A_{j,k} \in \mathbb{F}$ for all $j = 1, \dots, m$ and

all $k = 1, \dots, n$. For $c_1, \dots, c_m \in \mathbb{F}$, consider the system of linear equations

$$\begin{aligned} \sum_{k=1}^{n} A_{1,k} x_k &= c_1 \\ &\vdots \\ \sum_{k=1}^{n} A_{m,k} x_k &= c_m. \end{aligned} \tag{3.27}$$

With this notation, the question here is whether there is some choice of the constant terms $c_1, \dots, c_m \in \mathbb{F}$ such that no solution exists to the system above.

Define $T: \mathbb{F}^n \to \mathbb{F}^m$ as in 3.25. The equation $T(x_1, \dots, x_n) = (c_1, \dots, c_m)$ is the same as the system of equations 3.27. Thus we want to know if $\operatorname{im} T \neq \mathbb{F}^m$. Hence we can rephrase our question about not having a solution for some choice of $c_1, \dots, c_m \in \mathbb{F}$ as follows: What condition ensures that T is not surjective? The next result gives one such condition.

The results 3.26 and 3.28, which compare the number of variables and the number of equations, can also be proved using Gaussian elimination. The abstract approach taken here seems to provide cleaner proofs.

System of linear equations with more equations than variables **Result 3.28**

A system of linear equations with more equations than variables has no solution for some choice of the constant terms.

Proof. Use the notation from the discussion above. Thus T is a linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$, and we have a system of m equations with n variables $x_1, \dots, x_n$; see 3.27. If $n < m$, then 3.24 implies that T is not surjective. As discussed above, this shows that if we have more equations than variables in a system of linear equations, then there is no solution for some choice of the constant terms. □

Example of the result above: a system of five linear equations with four variables has no solution for some choice of the constant terms.

Exercises

3B.1 Give an example of a linear map T with $\dim \ker T = 3$ and $\dim \operatorname{im} T = 2$.

3B.2 Suppose $S, T \in \operatorname{End}(V)$ are such that $\operatorname{im} S \subseteq \ker T$. Prove that $(ST)^2 = 0$.

3B.3 Suppose $v_1, \dots, v_m$ is a list of vectors in V. Define $T \in \operatorname{Hom}(\mathbb{F}^m, V)$ by

$$T(z_1, \dots, z_m) = z_1 v_1 + \cdots + z_m v_m.$$

(a) What property of T corresponds to $v_1, \dots, v_m$ spanning V?

(b) What property of T corresponds to the list $v_1, \dots, v_m$ being linearly independent?

3B.4 Show that $\{T \in \text{Hom}(\mathbb{R}^5, \mathbb{R}^4) : \dim \ker T > 2\}$ is not a subspace of $\text{Hom}(\mathbb{R}^5, \mathbb{R}^4)$.

3B.5 Give an example of $T \in \text{End}(\mathbb{R}^4)$ such that $\text{im}\, T = \ker T$.

3B.6 Prove that there does not exist $T \in \text{End}(\mathbb{R}^5)$ such that $\text{im}\, T = \ker T$.

3B.7 Suppose V and W are finite-dimensional with $2 \le \dim V \le \dim W$. Show that $\{T \in \text{Hom}(V, W) : T \text{ is not injective}\}$ is not a subspace of $\text{Hom}(V, W)$.

3B.8 Suppose V and W are finite-dimensional with $\dim V \ge \dim W \ge 2$. Show that $\{T \in \text{Hom}(V, W) : T \text{ is not surjective}\}$ is not a subspace of $\text{Hom}(V, W)$.

3B.9 Suppose $T \in \text{Hom}(V, W)$ is injective and $v_1, \dots, v_n$ is linearly independent in V. Prove that $Tv_1, \dots, Tv_n$ is linearly independent in W.

3B.10 Suppose $v_1, \dots, v_n$ spans V and $T \in \text{Hom}(V, W)$. Show that $Tv_1, \dots, Tv_n$ spans $\text{im}\, T$.

3B.11 Suppose that V is finite-dimensional and that $T \in \text{Hom}(V, W)$. Prove that there exists a subspace U of V such that

$$U \cap \ker T = \{0\} \quad \text{and} \quad \text{im}\, T = \{Tu : u \in U\}.$$

3B.12 Suppose T is a linear map from $\mathbb{F}^4$ to $\mathbb{F}^2$ such that

$$\ker T = \{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 : x_1 = 5x_2 \text{ and } x_3 = 7x_4\}.$$

Prove that T is surjective.

3B.13 Suppose U is a three-dimensional subspace of $\mathbb{R}^8$ and that T is a linear map from $\mathbb{R}^8$ to $\mathbb{R}^5$ such that $\ker T = U$. Prove that T is surjective.

3B.14 Prove that there does not exist a linear map from $\mathbb{F}^5$ to $\mathbb{F}^2$ whose null space equals $\{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{F}^5 : x_1 = 3x_2 \text{ and } x_3 = x_4 = x_5\}$.

3B.15 Suppose there exists a linear map on V whose null space and range are both finite-dimensional. Prove that V is finite-dimensional.

3B.16 Suppose V and W are both finite-dimensional. Prove that there exists an injective linear map from V to W if and only if $\dim V \le \dim W$.

3B.17 Suppose V and W are both finite-dimensional. Prove that there exists a surjective linear map from V onto W if and only if $\dim V \geq \dim W$.

3B.18 Suppose V and W are finite-dimensional and that U is a subspace of V. Prove that there exists $T \in \operatorname{Hom}(V, W)$ such that $\ker T = U$ if and only if $\dim U \geq \dim V - \dim W$.

3B.19 Suppose W is finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Prove that T is injective if and only if there exists $S \in \operatorname{Hom}(W, V)$ such that ST is the identity operator on V.

3B.20 Suppose W is finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Prove that T is surjective if and only if there exists $S \in \operatorname{Hom}(W, V)$ such that TS is the identity operator on W.

3B.21 Suppose V is finite-dimensional, $T \in \operatorname{Hom}(V, W)$, and U is a subspace of W. Prove that $\{v \in V : Tv \in U\}$ is a subspace of V and

$$\dim\{v \in V : Tv \in U\} = \dim \ker T + \dim(U \cap \operatorname{im} T).$$

3B.22 Suppose U and V are finite-dimensional vector spaces and $S \in \operatorname{Hom}(V, W)$ and $T \in \operatorname{Hom}(U, V)$. Prove that

$$\dim \ker ST \leq \dim \ker S + \dim \ker T.$$

3B.23 Suppose U and V are finite-dimensional vector spaces and $S \in \operatorname{Hom}(V, W)$ and $T \in \operatorname{Hom}(U, V)$. Prove that

$$\dim \operatorname{im} ST \leq \min\{\dim \operatorname{im} S, \dim \operatorname{im} T\}.$$

3B.24

(a) Suppose $\dim V = 5$ and $S, T \in \operatorname{End}(V)$ are such that $ST = 0$. Prove that $\dim \operatorname{im} TS \leq 2$.

(b) Give an example of $S, T \in \operatorname{End}(\mathbb{F}^5)$ with $ST = 0$ and $\dim \operatorname{im} TS = 2$.

3B.25 Suppose that W is finite-dimensional and $S, T \in \operatorname{Hom}(V, W)$. Prove that $\ker S \subseteq \ker T$ if and only if there exists $E \in \operatorname{End}(W)$ such that $T = ES$.

3B.26 Suppose V is finite-dimensional and $S, T \in \text{Hom}(V, W)$. Prove that $\text{im}\, S \subseteq \text{im}\, T$ if and only if there exists $E \in \text{End}(V)$ such that $S = TE$.

3B.27 Suppose $P \in \text{End}(V)$ and $P^2 = P$. Prove that $V = \ker P \oplus \text{im}\, P$.

3B.28 Suppose $D \in \text{End}(\mathbb{R}[X])$ is such that $\deg Dp = (\deg p) - 1$ for every nonconstant polynomial $p \in \mathbb{R}[X]$. 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'.

3B.29 Suppose $p \in \mathbb{R}[X]$. Prove that there exists a polynomial $q \in \mathbb{R}[X]$ such that $5q'' + 3q' = p$.

This exercise can be done without linear algebra, but it's more fun to do it using linear algebra.

3B.30 Suppose $\varphi \in \text{Hom}(V, \mathbb{F})$ and $\varphi \neq 0$. Suppose $u \in V$ is not in $\ker \varphi$. Prove that

$$V = \ker \varphi \oplus \{au : a \in \mathbb{F}\}.$$

3B.31 Suppose V is finite-dimensional, X is a subspace of V, and Y is a finite-dimensional subspace of W. Prove that there exists $T \in \text{Hom}(V, W)$ such that $\ker T = X$ and $\text{im}\, T = Y$ if and only if $\dim X + \dim Y = \dim V$.

3B.32 Suppose V is finite-dimensional with $\dim V > 1$. Show that if $\varphi\colon \text{End}(V) \to \mathbb{F}$ is a linear map such that $\varphi(ST) = \varphi(S)\varphi(T)$ for all $S, T \in \text{End}(V)$, then $\varphi = 0$.

Hint: The description of the two-sided ideals of $\text{End}(V)$ *given by Exercise 17 in Section 3A might be useful.*

3B.33 Suppose that V and W are real vector spaces and $T \in \text{Hom}(V, W)$. Define $T_{\mathbb{C}}\colon V_{\mathbb{C}} \to W_{\mathbb{C}}$ by

$$T_{\mathbb{C}}(u + iv) = Tu + iTv$$

for all $u, v \in V$.

(a) Show that $T_{\mathbb{C}}$ is a (complex) linear map from $V_{\mathbb{C}}$ to $W_{\mathbb{C}}$.

(b) Show that $T_{\mathbb{C}}$ is injective if and only if T is injective.

(c) Show that $\text{im}\, T_{\mathbb{C}} = W_{\mathbb{C}}$ if and only if $\text{im}\, T = W$.

See Exercise 8 in Section 1B for the definition of the complexification $V_{\mathbb{C}}$. The linear map $T_{\mathbb{C}}$ is called the ***complexification*** *of the linear map T.*

3C 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 the linear map lemma (3.4). As we will soon see, matrices provide an efficient method of recording the values of the Tv_k's in terms of a basis of W.

Matrix, $A_{j,k}$ **Definition 3.29**

Suppose m and n are nonnegative integers. An *m-by-n matrix A* is a rectangular array of elements of $\mathbb{F}$ with m rows and n columns:

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}.$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A.

When dealing with matrices, the first index refers to the row number; the second index refers to the column number.

Example 3.30 **$A_{j,k}$ equals entry in row j, column k of A**

Suppose $A = \begin{pmatrix} 8 & 4 & 5-3i \\ 1 & 9 & 7 \end{pmatrix}$. Thus $A_{2,3}$ refers to the entry in the second row, third column of A, which means that $A_{2,3} = 7$.

Now we come to the key definition in this section.

Matrix of a linear map, $\mathrm{Mat}(T)$ **Definition 3.31**

Suppose $T \in \mathrm{Hom}(V, W)$ 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 $\mathrm{Mat}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are not clear from the context, then the notation $\mathrm{Mat}\big(T, (v_1, \dots, v_n), (w_1, \dots, w_m)\big)$ is used.

The matrix $\mathrm{Mat}(T)$ of a linear map $T \in \mathrm{Hom}(V, W)$ 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 $\mathrm{Mat}(T)$ 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:

$$\operatorname{Mat}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & \cdots & v_k & \cdots & v_n \end{array} \\ \begin{pmatrix} & A_{1,k} & \\ & \vdots & \\ & A_{m,k} & \end{pmatrix} \end{array}.$$

In the matrix above only the kth 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 $\operatorname{Mat}(T)$ by multiplying each entry in the kth column by the corresponding w_j from the left column, and then adding up the resulting vectors.

If T is a linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$, then unless stated otherwise, assume the bases in question are the standard ones (where the kth basis vector is 1 in the kth slot and 0 in all other slots). If you think of elements of $\mathbb{F}^m$ as columns of m numbers, then you can think of the kth column of $\operatorname{Mat}(T)$ as T applied to the kth standard basis vector.

The kth column of $\operatorname{Mat}(T)$ 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 an n-dimensional vector space to an m-dimensional vector space, then $\operatorname{Mat}(T)$ is an m-by-n matrix.

Example 3.32 **The matrix of a linear map from $\mathbb{F}^2$ to $\mathbb{F}^3$**

Suppose $T \in \operatorname{Hom}(\mathbb{F}^2, \mathbb{F}^3)$ is defined by

$$T(x, y) = (x + 3y,\ 2x + 5y,\ 7x + 9y).$$

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:

$$\operatorname{Mat}(T) = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ 7 & 9 \end{pmatrix}.$$

When working with $\mathbb{F}[X]_{\le m}$, use the standard basis $1, x, x^2, \dots, x^m$ unless the context indicates otherwise.

Example 3.33 **Matrix of the differentiation map from $\mathbb{R}[X]_{\le 3}$ to $\mathbb{R}[X]_{\le 2}$**

Suppose $D \in \operatorname{Hom}\big(\mathbb{R}[X]_{\le 3}, \mathbb{R}[X]_{\le 2}\big)$ is the differentiation map defined by $Dp = p'$. Because $(x^n)' = nx^{n-1}$, the matrix of D with respect to the standard bases is the 3-by-4 matrix below:

$$\operatorname{Mat}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Addition and Scalar Multiplication of Matrices

For the rest of this section, assume that U, 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).

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 yet make sense because although we have defined the sum of two linear maps, we have not defined the sum of two matrices. Fortunately, the natural definition of the sum of two matrices has the right properties. Specifically, we make the following definition.

Matrix addition — Definition 3.34

The *sum of two matrices of the same size* is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \cdots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \cdots & C_{m,n} \end{pmatrix} = \begin{pmatrix} A_{1,1} + C_{1,1} & \cdots & A_{1,n} + C_{1,n} \\ \vdots & & \vdots \\ A_{m,1} + C_{m,1} & \cdots & A_{m,n} + C_{m,n} \end{pmatrix}.$$

In the next result, the assumption is that the same bases are used for all three linear maps $S + T$, S, and T.

Matrix of the sum of linear maps — Result 3.35

Suppose $S, T \in \text{Hom}(V, W)$. Then $\text{Mat}(S + T) = \text{Mat}(S) + \text{Mat}(T)$.

The verification of the result above follows from the definitions and 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 yet make sense because we have not defined scalar multiplication on matrices. Fortunately, the natural definition again has the right properties.

Scalar multiplication of a matrix — Definition 3.36

The product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \cdots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \cdots & \lambda A_{m,n} \end{pmatrix}.$$

Example 3.37 **Addition and scalar multiplication of matrices**

$$2\begin{pmatrix} 3 & 1 \\ -1 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 2 \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ -2 & 10 \end{pmatrix} + \begin{pmatrix} 4 & 2 \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} 10 & 4 \\ -1 & 16 \end{pmatrix}$$

In the next result, the assumption is that the same bases are used for both the linear maps λT and T.

The matrix of a scalar times a linear map **Result 3.38**

Suppose $\lambda \in \mathbb{F}$ and $T \in \text{Hom}(V, W)$. Then $\text{Mat}(\lambda T) = \lambda \, \text{Mat}(T)$.

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. First we introduce a bit of notation so that this new vector space has a name, and then we find the dimension of this new vector space.

$\mathbb{F}^{m,n}$ **Notation 3.39**

For m and n positive integers, the set of all m-by-n matrices with entries in $\mathbb{F}$ is denoted by $\mathbb{F}^{m,n}$.

$\dim \mathbb{F}^{m,n} = mn$ **Result 3.40**

Suppose m and n are positive integers. With addition and scalar multiplication defined as above, $\mathbb{F}^{m,n}$ is a vector space of dimension mn.

Proof. The verification that $\mathbb{F}^{m,n}$ is a vector space is left to the reader. Note that the additive identity of $\mathbb{F}^{m,n}$ is the m-by-n matrix all of whose entries equal 0.

The reader should also verify that the list of distinct m-by-n matrices that have 0 in all entries except for a 1 in one entry is a basis of $\mathbb{F}^{m,n}$. There are mn such matrices, so the dimension of $\mathbb{F}^{m,n}$ equals mn. □

Matrix Multiplication

Suppose, as previously, that $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Suppose also that $u_1, \ldots, u_p$ 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 $\text{Mat}(ST)$ equal $\text{Mat}(S)\,\text{Mat}(T)$? 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 $\text{Mat}(S) = A$ and $\text{Mat}(T) = B$. For $1 \le k \le p$, we have

$$\begin{aligned}(ST)u_k &= S\Big(\sum_{r=1}^{n} B_{r,k} v_r\Big) \\ &= \sum_{r=1}^{n} B_{r,k} S v_r \\ &= \sum_{r=1}^{n} B_{r,k} \sum_{j=1}^{m} A_{j,r} w_j \\ &= \sum_{j=1}^{m} \Big(\sum_{r=1}^{n} A_{j,r} B_{r,k}\Big) w_j .\end{aligned}$$

Thus $\text{Mat}(ST)$ is the m-by-p matrix whose entry in row j, column k, equals

$$\sum_{r=1}^{n} A_{j,r} B_{r,k}.$$

Now we see how to define matrix multiplication so that the desired equation $\text{Mat}(ST) = \text{Mat}(S)\,\text{Mat}(T)$ holds.

Matrix multiplication **Definition 3.41**

Suppose A is an m-by-n matrix and B is an n-by-p matrix. Then AB is defined to be the m-by-p matrix whose entry in row j, column k, is given by the equation

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}.$$

Thus the entry in row j, column k, of AB is computed by taking row j of A and column k of B, 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 this motivation for it.

Example 3.42 **Matrix multiplication**

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—AB is not necessarily equal to BA even if both products are defined (see Exercise 10). Matrix multiplication is distributive and associative (see Exercises 11 and 12).

In the next result, we assume that the same basis of V is used in considering $T \in \text{Hom}(U, V)$ and $S \in \text{Hom}(V, W)$, the same basis of W is used in considering

$S \in \text{Hom}(V, W)$ and $ST \in \text{Hom}(U, W)$, and the same basis of U is used in considering $T \in \text{Hom}(U, V)$ and $ST \in \text{Hom}(U, W)$.

Matrix of product of linear maps — Result 3.43

If $T \in \text{Hom}(U, V)$ and $S \in \text{Hom}(V, W)$, then $\text{Mat}(ST) = \text{Mat}(S)\,\text{Mat}(T)$.

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.

$A_{j,\cdot}$, $A_{\cdot,k}$ — Notation 3.44

Suppose A is an m-by-n matrix.

- If $1 \le j \le m$, then $A_{j,\cdot}$ denotes the 1-by-n matrix consisting of row j of A.
- If $1 \le k \le n$, then $A_{\cdot,k}$ denotes the m-by-1 matrix consisting of column k of A.

Example 3.45 $A_{j,\cdot}$ **equals** j^{th} **row of** A **and** $A_{\cdot,k}$ **equals** k^{th} **column of** A

The notation $A_{2,\cdot}$ denotes the second row of A and $A_{\cdot,2}$ denotes the second column of A. Thus if $A = \begin{pmatrix} 8 & 4 & 5 \\ 1 & 9 & 7 \end{pmatrix}$, then

$$A_{2,\cdot} = \begin{pmatrix} 1 & 9 & 7 \end{pmatrix} \quad \text{and} \quad A_{\cdot,2} = \begin{pmatrix} 4 \\ 9 \end{pmatrix}.$$

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. For 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 $\begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 6 \\ 2 \end{pmatrix} = 26$.

The next result uses the convention discussed in the paragraph above to give another way to think of matrix multiplication. For example, the next result and the calculation in the paragraph above explain why the entry in row 2, column 1, of the product in Example 3.42 equals 26.

Result 3.46 **Entry of matrix product equals row times column**

Suppose A is an m-by-n matrix and B is an n-by-p matrix. Then

$$(AB)_{j,k} = A_{j,\cdot}B_{\cdot,k}$$

if $1 \le j \le m$ and $1 \le k \le p$. In other words, the entry in row j, column k, of AB equals (row j of A) times (column k of B).

Proof. Suppose $1 \le j \le m$ and $1 \le k \le p$. The definition of matrix multiplication states that

$$(AB)_{j,k} = A_{j,1}B_{1,k} + \cdots + A_{j,n}B_{n,k}. \tag{3.47}$$

The definition of matrix multiplication also implies that the product of the 1-by-n matrix $A_{j,\cdot}$ and the n-by-1 matrix $B_{\cdot,k}$ is the 1-by-1 matrix whose entry is the number on the right side of the equation above. Thus the entry in row j, column k, of AB equals (row j of A) times (column k of B). □

The next result gives yet another way to think of matrix multiplication. In the result below, $(AB)_{\cdot,k}$ is column k of the m-by-p matrix AB. Thus $(AB)_{\cdot,k}$ is an m-by-1 matrix. Also, $AB_{\cdot,k}$ is an m-by-1 matrix because it is the product of an m-by-n matrix and an n-by-1 matrix. Thus the two sides of the equation in the result below have the same size, making it reasonable that they might be equal.

Result 3.48 **Column of matrix product equals matrix times column**

Suppose A is an m-by-n matrix and B is an n-by-p matrix. Then

$$(AB)_{\cdot,k} = AB_{\cdot,k}$$

if $1 \le k \le p$. In other words, column k of AB equals A times column k of B.

Proof. As discussed above, $(AB)_{\cdot,k}$ and $AB_{\cdot,k}$ are both m-by-1 matrices. If $1 \le j \le m$, then the entry in row j of $(AB)_{\cdot,k}$ is the left side of 3.47 and the entry in row j of $AB_{\cdot,k}$ is the right side of 3.47. Thus $(AB)_{\cdot,k} = AB_{\cdot,k}$. □

Our next result will give another way of thinking about the product of an m-by-n matrix and an n-by-1 matrix, motivated by the next example.

Example 3.49 **Product of a 3-by-2 matrix and a 2-by-1 matrix**

Use our definitions and basic arithmetic to verify that

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 19 \\ 31 \end{pmatrix} = 5 \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}.$$

Thus in this example, 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 (5 and 1) that multiply the columns coming from the 2-by-1 matrix.

The next result generalizes the example above.

Linear combination of columns **Result 3.50**

Suppose A is an m-by-n matrix and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ is an n-by-1 matrix. Then

$$Ab = b_1 A_{\cdot,1} + \cdots + b_n A_{\cdot,n}.$$

In other words, Ab is a linear combination of the columns of A, with the scalars that multiply the columns coming from b.

Proof. If $k \in \{1, \ldots, m\}$, then the definition of matrix multiplication implies that the entry in row k of the m-by-1 matrix Ab is

$$A_{k,1} b_1 + \cdots + A_{k,n} b_n.$$

The entry in row k of $b_1 A_{\cdot,1} + \cdots + b_n A_{\cdot,n}$ also equals the number displayed above.

Because Ab and $b_1 A_{\cdot,1} + \cdots + b_n A_{\cdot,n}$ have the same entry in row k for each $k \in \{1, \ldots, m\}$, we conclude that $Ab = b_1 A_{\cdot,1} + \cdots + b_n A_{\cdot,n}$. □

Our two previous results focus on the columns of a matrix. Analogous results hold for the rows of a matrix. Specifically, see Exercises 8 and 9, which can be proved using appropriate modifications of the proofs of 3.48 and 3.50.

The next result is the main tool used in the next subsection to prove the column–row factorization (3.56) and to prove that the column rank of a matrix equals the row rank (3.57). To be consistent with the notation often used with the column–row factorization, including in the next subsection, the matrices in the next result are called C and R instead of A and B.

Matrix multiplication as linear combinations of columns or rows **Result 3.51**

Suppose C is an m-by-c matrix and R is a c-by-n matrix.

(a) If $k \in \{1, \ldots, n\}$, then column k of CR is a linear combination of the columns of C, with the coefficients of this linear combination coming from column k of R.

(b) If $j \in \{1, \ldots, m\}$, then row j of CR is a linear combination of the rows of R, with the coefficients of this linear combination coming from row j of C.

Proof. Suppose $k \in \{1, \ldots, n\}$. Then column k of CR equals $CR_{\cdot,k}$ (by 3.48), which equals the linear combination of the columns of C with coefficients coming from $R_{\cdot,k}$ (by 3.50). Thus (a) holds.

To prove (b), follow the pattern of the proof of (a) but use rows instead of columns and use Exercises 8 and 9 instead of 3.48 and 3.50. □

Column–Row Factorization and Rank of a Matrix

We begin by defining two nonnegative integers associated with each matrix.

Column rank, row rank — Definition 3.52

Suppose A is an m-by-n matrix with entries in $\mathbb{F}$.

- The *column rank* of A is the dimension of the span of the columns of A in $\mathbb{F}^{m,1}$.
- The *row rank* of A is the dimension of the span of the rows of A in $\mathbb{F}^{1,n}$.

If A is an m-by-n matrix, then the column rank of A is at most n (because A has n columns) and the column rank of A is also at most m (because $\dim \mathbb{F}^{m,1} = m$). Similarly, the row rank of A is also at most $\min\{m, n\}$.

Example 3.53 **Column rank and row rank of a 2-by-4 matrix**

Suppose

$$A = \begin{pmatrix} 4 & 7 & 1 & 8 \\ 3 & 5 & 2 & 9 \end{pmatrix}.$$

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 $\mathbb{F}^{2,1}$. Neither of the first two vectors listed above in $\mathbb{F}^{2,1}$ is a scalar multiple of the other. Thus the span of this list of length four has dimension at least two. The span of this list of vectors in $\mathbb{F}^{2,1}$ cannot have dimension larger than two because $\dim \mathbb{F}^{2,1} = 2$. Thus the span of this list has dimension two, which means that the column rank of A is two.

The row rank of A is the dimension of

$$\operatorname{span}\Big(\begin{pmatrix} 4 & 7 & 1 & 8 \end{pmatrix}, \begin{pmatrix} 3 & 5 & 2 & 9 \end{pmatrix}\Big)$$

in $\mathbb{F}^{1,4}$. Neither of the two vectors listed above in $\mathbb{F}^{1,4}$ is a scalar multiple of the other. Thus the span of this list of length two has dimension two, which means that the row rank of A is two.

We now define the transpose of a matrix.

Transpose, A^t — Definition 3.54

The *transpose* of a matrix A, denoted by A^t, is the matrix obtained from A by interchanging rows and columns. Specifically, if A is an m-by-n matrix, then A^t

Definition 3.54 *continued*

is the n-by-m matrix whose entries are given by the equation

$$(A^t)_{k,j} = A_{j,k}.$$

Example 3.55 **Transpose of a matrix**

If $A = \begin{pmatrix} 5 & -7 \\ 3 & 8 \\ -4 & 2 \end{pmatrix}$, then $A^t = \begin{pmatrix} 5 & 3 & -4 \\ -7 & 8 & 2 \end{pmatrix}$.

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: $(A + B)^t = A^t + B^t$, $(\lambda A)^t = \lambda A^t$, and $(AC)^t = C^t A^t$ for all m-by-n matrices A, B, all $\lambda \in \mathbb{F}$, and all n-by-p matrices C (see Exercises 14 and 15).

The next result will be the main tool used to prove that the column rank equals the row rank (see 3.57).

Column–row factorization **Result 3.56**

Suppose A is an m-by-n matrix with entries in $\mathbb{F}$ and column rank $c \geq 1$. Then there exist an m-by-c matrix C and a c-by-n matrix R, both with entries in $\mathbb{F}$, such that $A = CR$.

Proof. Each column of A is an m-by-1 matrix. The list $A_{\cdot,1}, \ldots, A_{\cdot,n}$ of columns of A can be reduced to a basis of the span of the columns of A (by 2.30). This basis has length c, by the definition of the column rank. The c columns in this basis can be put together to form an m-by-c matrix C.

If $k \in \{1, \ldots, n\}$, then column k of A is a linear combination of the columns of C. Make the coefficients of this linear combination into column k of a c-by-n matrix that we call R. Then $A = CR$, as follows from 3.51(a). □

In Example 3.53, the column rank and row rank turned out to equal each other. The next result states that this happens for all matrices.

Column rank equals row rank **Result 3.57**

Suppose $A \in \mathbb{F}^{m,n}$. Then the column rank of A equals the row rank of A.

Proof. Let c denote the column rank of A. If $c = 0$, then $A = 0$ and hence the row rank of A also equals 0. Thus we can assume that $c \geq 1$.

Let $A = CR$ be the column–row factorization of A given by 3.56, where C is an m-by-c matrix and R is a c-by-n matrix. Then 3.51(b) tells us that every row of A is a linear combination of the rows of R. Because R has c rows, this implies that the row rank of A is less than or equal to the column rank c of A.

To prove the inequality in the other direction, apply the previous paragraph result to A^t, getting

$$\begin{aligned}\text{column rank of } A &= \text{row rank of } A^t\\ &\le \text{column rank of } A^t\\ &= \text{row rank of } A.\end{aligned}$$

Thus the column rank of A equals the row rank of A. □

Because the column rank equals the row rank, the last result allows us to dispense with the terms "column rank" and "row rank" and just use the simpler term "rank".

Rank — Definition 3.58

The *rank* of a matrix $A \in \mathbb{F}^{m,n}$ is the column rank of A.

See 3.133 and Exercise 8 in Section 7A for alternative proofs that the column rank equals the row rank.

Exercises

3C.1 Suppose $T \in \text{Hom}(V, W)$. Show that with respect to each choice of bases of V and W, the matrix of T has at least $\dim \text{im}\, T$ nonzero entries.

3C.2 Suppose $T \in \text{Hom}(V, W)$, where V and W are finite-dimensional and nonzero. Prove that $\dim \text{im}\, T = 1$ 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 $\text{Mat}(T)$ equal 1.

3C.3 Suppose $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W.

(a) Show that if $S, T \in \text{Hom}(V, W)$, then $\text{Mat}(S + T) = \text{Mat}(S) + \text{Mat}(T)$.

(b) Show that if $\lambda \in \mathbb{F}$ and $T \in \text{Hom}(V, W)$, then $\text{Mat}(\lambda T) = \lambda\, \text{Mat}(T)$.

This exercise asks you to verify 3.35 and 3.38.

3C.4 Suppose that $D \in \text{Hom}\big(\mathbb{R}[X]_{\le 3}, \mathbb{R}[X]_{\le 2}\big)$ is the differentiation map defined by $Dp = p'$. Find a basis of $\mathbb{R}[X]_{\le 3}$ and a basis of $\mathbb{R}[X]_{\le 2}$ 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 with Example 3.33. The next exercise generalizes this exercise.

3C.5 Suppose V and W are finite-dimensional and $T \in \text{Hom}(V, W)$. Prove that there exist a basis of V and a basis of W such that with respect to these bases, all entries of $\text{Mat}(T)$ are 0 except that the entries in row k, column k, equal 1 if $1 \le k \le \dim \text{im}\, T$.

3C.6 Suppose $v_1, \dots, v_m$ is a basis of V and W is finite-dimensional. Suppose $T \in \text{Hom}(V, W)$. Prove that there exists a basis $w_1, \dots, w_n$ of W such that all entries in the first column of $\text{Mat}(T)$ [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 5, you are given the basis of V instead of being able to choose a basis of V.

3C.7 Suppose $w_1, \dots, w_n$ is a basis of W and V is finite-dimensional. Suppose $T \in \text{Hom}(V, W)$. Prove that there exists a basis $v_1, \dots, v_m$ of V such that all entries in the first row of $\text{Mat}(T)$ [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 5, you are given the basis of W instead of being able to choose a basis of W.

3C.8 Suppose A is an m-by-n matrix and B is an n-by-p matrix. Prove that

$$(AB)_{j,\cdot} = A_{j,\cdot}B$$

for each $1 \le j \le m$. In other words, show that row j of AB equals (row j of A) times B.

This exercise gives the row version of 3.48.

3C.9 Suppose $a = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$ is a 1-by-n matrix and B is an n-by-p matrix. Prove that

$$aB = a_1 B_{1,\cdot} + \cdots + a_n B_{n,\cdot}.$$

In other words, show that aB is a linear combination of the rows of B, with the scalars that multiply the rows coming from a.

This exercise gives the row version of 3.50.

3C.10 Give an example of 2-by-2 matrices A and B such that $AB \ne BA$.

3C.11 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. Explain why $AB + AC$ and $DF + EF$ both make sense and prove that

$$A(B + C) = AB + AC \quad \text{and} \quad (D + E)F = DF + EF.$$

3C.12 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. Explain why $A(BC)$ makes sense and prove that

$$(AB)C = A(BC).$$

Try to find a clean proof that illustrates the following quote from Emil Artin: "It is my experience that proofs involving matrices can be shortened by 50% if one throws the matrices out."

3C.13 Suppose A is an n-by-n matrix and $1 \le j, k \le n$. Show that the entry in row j, column k, of A^3 (which is defined to mean AAA) is

$$\sum_{p=1}^{n} \sum_{r=1}^{n} A_{j,p} A_{p,r} A_{r,k}.$$

3C.14 Suppose m and n are positive integers. Prove that the function $A \mapsto A^t$ is a linear map from $\mathbb{F}^{m,n}$ to $\mathbb{F}^{n,m}$.

3C.15 Prove that if A is an m-by-n matrix and C is an n-by-p matrix, then

$$(AC)^t = C^t A^t.$$

This exercise shows that the transpose of the product of two matrices is the product of the transposes in the opposite order.

3C.16 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 $(c_1, \ldots, c_m) \in \mathbb{F}^m$ and $(d_1, \ldots, d_n) \in \mathbb{F}^n$ such that $A_{j,k} = c_j d_k$ for every $j = 1, \ldots, m$ and every $k = 1, \ldots, n$.

3C.17 Suppose $T \in \text{End}(V)$, and $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Prove that the following are equivalent.

(a) T is injective.

(b) The columns of $\text{Mat}(T)$ are linearly independent in $\mathbb{F}^{n,1}$.

(c) The columns of $\text{Mat}(T)$ span $\mathbb{F}^{n,1}$.

(d) The rows of $\text{Mat}(T)$ span $\mathbb{F}^{1,n}$.

(e) The rows of $\text{Mat}(T)$ are linearly independent in $\mathbb{F}^{1,n}$.

Here $\text{Mat}(T)$ means $\text{Mat}\big(T, (u_1, \ldots, u_n), (v_1, \ldots, v_n)\big)$.

3D Invertibility and Isomorphisms

Invertible Linear Maps

We begin this section by defining the notions of invertible and inverse in the context of linear maps.

Invertible, inverse — Definition 3.59

- A linear map $T \in \text{Hom}(V, W)$ is called *invertible* if there exists a linear map $S \in \text{Hom}(W, V)$ such that ST equals the identity operator on V and TS equals the identity operator on W.
- A linear map $S \in \text{Hom}(W, V)$ satisfying $ST = I$ and $TS = I$ is called an *inverse* of T (note that the first I is the identity operator on V and the second I is the identity operator on W).

The definition above mentions "an inverse". However, the next result shows that we can change this terminology to "the inverse".

Inverse is unique — Result 3.60

An invertible linear map has a unique inverse.

Proof. Suppose $T \in \text{Hom}(V, W)$ 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.

T^{-1} — Notation 3.61

If T is invertible, then its inverse is denoted by T^{-1}. In other words, if $T \in \text{Hom}(V, W)$ is invertible, then T^{-1} is the unique element of $\text{Hom}(W, V)$ such that $T^{-1}T = I$ and $TT^{-1} = I$.

Example 3.62 **Inverse of a linear map from $\mathbb{R}^3$ to $\mathbb{R}^3$**

Suppose $T \in \text{End}(\mathbb{R}^3)$ is defined by $T(x, y, z) = (-y, x, 4z)$. Thus T is a counterclockwise rotation by 90° in the xy-plane and a stretch by a factor of 4 in the direction of the z-axis.

Hence the inverse map $T^{-1} \in \text{End}(\mathbb{R}^3)$ is the clockwise rotation by 90° in the xy-plane and a stretch by a factor of $1/4$ in the direction of the z-axis:

$$T^{-1}(x, y, z) = (y, -x, 1/4z).$$

The next result shows that a linear map is invertible if and only if it is one-to-one and onto.

Invertibility $\iff$ injectivity and surjectivity **Result 3.63**

A linear map is invertible if and only if it is injective and surjective.

Proof. Suppose $T \in \text{Hom}(V, W)$. 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 $\text{im}\, T = W$. 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 $S(w)$ to be the unique element of V such that $T\big(S(w)\big) = w$ (the existence and uniqueness of such an element follow from the surjectivity and injectivity of T). The definition of S implies that $T \circ S$ equals the identity operator on W.

To prove that $S \circ T$ equals the identity operator 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 operator 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

$$T\big(S(w_1) + S(w_2)\big) = T\big(S(w_1)\big) + T\big(S(w_2)\big) = w_1 + w_2.$$

Thus $S(w_1) + S(w_2)$ is the unique element of V that T maps to $w_1 + w_2$. By the definition of S, this implies that $S(w_1 + w_2) = S(w_1) + S(w_2)$. Hence S satisfies the additive property required for linearity.

The proof of homogeneity is similar. Specifically, if $w \in W$ and $\lambda \in \mathbb{F}$, then

$$T\big(\lambda S(w)\big) = \lambda T\big(S(w)\big) = \lambda w.$$

Thus $\lambda S(w)$ is the unique element of V that T maps to λw. By the definition of S, this implies that $S(\lambda w) = \lambda S(w)$. Hence S is linear, as desired. □

For a linear map from a vector space to itself, 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 linear maps from Example 3.3.

Example 3.64 **Neither injectivity nor surjectivity implies invertibility**

Example 3.64 *continued*

- The multiplication by x^2 linear map from $\mathbb{R}[X]$ to $\mathbb{R}[X]$ (see 3.3) is injective but it is not invertible because it is not surjective (the polynomial 1 is not in the range).
- The backward shift linear map from $\mathbb{F}^\infty$ to $\mathbb{F}^\infty$ (see 3.3) is surjective but it is not invertible because it is not injective [the vector $(1, 0, 0, 0, \dots)$ is in the null space].

In view of the example above, the next result is remarkable—it states that for a linear map from a finite-dimensional vector space to a vector space of the same dimension, either injectivity or surjectivity alone implies the other condition. Note that the hypothesis below that $\dim V = \dim W$ is automatically satisfied in the important special case where V is finite-dimensional and $W = V$.

Result 3.65

Injectivity is equivalent to surjectivity (if $\dim V = \dim W < \infty$)
Suppose that V and W are finite-dimensional vector spaces, $\dim V = \dim W$, and $T \in \operatorname{Hom}(V, W)$. Then

$$T \text{ is invertible} \iff T \text{ is injective} \iff T \text{ is surjective.}$$

Proof. The fundamental theorem of linear maps (3.21) states that

$$\dim V = \dim \ker T + \dim \operatorname{im} T. \tag{3.66}$$

If T is injective (which by 3.15 is equivalent to the condition $\dim \ker T = 0$), then the equation above implies that

$$\dim \operatorname{im} T = \dim V - \dim \ker T = \dim V = \dim W,$$

which implies that T is surjective (by 2.39).

Conversely, if T is surjective, then 3.66 implies that

$$\dim \ker T = \dim V - \dim \operatorname{im} T = \dim V - \dim W = 0,$$

which implies that T is injective.

Thus we have shown that T is injective if and only if T is surjective. Thus if T is either injective or surjective, then T is both injective and surjective, which implies that T is invertible. Hence T is invertible if and only if T is injective if and only if T is surjective. □

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.

Example 3.67 **There exists a polynomial p such that $\big((x^2 + 5x + 7)p\big)'' = q$**

Example 3.67 *continued*

The linear map

$$p \mapsto \big((x^2 + 5x + 7)p\big)''$$

from $\mathbb{R}[X]$ to itself is injective, as you can show. Thus we are tempted to use 3.65 to show that this map is surjective. However, Example 3.64 shows that the magic of 3.65 does not apply to the infinite-dimensional vector space $\mathbb{R}[X]$. We will get around this problem by restricting attention to the finite-dimensional vector space $\mathbb{R}[X]_{\le m}$.

Suppose $q \in \mathbb{R}[X]$. There exists a nonnegative integer m such that $q \in \mathbb{R}[X]_{\le m}$. Define $T\colon \mathbb{R}[X]_{\le m} \to \mathbb{R}[X]_{\le m}$ 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 a linear map from $\mathbb{R}[X]_{\le m}$ to itself.

Every polynomial whose second derivative equals 0 is of the form $ax + b$, where $a, b \in \mathbb{R}$. Thus $\ker T = \{0\}$. Hence T is injective.

Thus T is surjective (by 3.65), which means that there exists a polynomial $p \in \mathbb{R}[X]_{\le m}$ such that $\big((x^2 + 5x + 7)p\big)'' = q$, as claimed in the title of this example.

Exercise 35 in Section 6A gives a similar but more spectacular example of using 3.65.

The hypothesis in the result below that $\dim V = \dim W$ holds in the important special case in which V is finite-dimensional and $W = V$. Thus in that case, the equation $ST = I$ implies that $ST = TS$, even though we do not have multiplicative commutativity of arbitrary linear maps from V to V.

Result 3.68

$ST = I \iff TS = I$ (on vector spaces of the same dimension)
Suppose V and W are finite-dimensional vector spaces of the same dimension, $S \in \text{Hom}(W, V)$, and $T \in \text{Hom}(V, W)$. Then $ST = I$ if and only if $TS = I$.

Proof. First suppose $ST = I$. If $v \in V$ and $Tv = 0$, then

$$v = Iv = (ST)v = S(Tv) = S(0) = 0.$$

Thus T is injective (by 3.15). Because V and W have the same dimension, this implies that T is invertible (by 3.65).

Now multiply both sides of the equation $ST = I$ by T^{-1} on the right, getting

$$S = T^{-1}.$$

Thus $TS = TT^{-1} = I$, as desired.

To prove the implication in the other direction, simply reverse the roles of S and T (and V and W) in the direction we have already proved, showing that if $TS = I$, then $ST = I$. □

Isomorphic Vector Spaces

The next definition captures the idea of two vector spaces that are essentially the same, except for the names of their elements.

Isomorphism, isomorphic — Definition 3.69

- 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.

It can be difficult to determine whether two mathematical structures (such as groups or topological spaces) are essentially the same, differing only in the names of the elements of underlying sets. However, the next result shows that we need to look at only a single number (the dimension) to determine whether two vector spaces are isomorphic.

Dimension shows whether vector spaces are isomorphic — Result 3.70

Two finite-dimensional vector spaces over $\mathbb{F}$ 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 $\ker T = \{0\}$ and $\operatorname{im} T = W$. Thus

$$\dim \ker T = 0 \quad \text{and} \quad \dim \operatorname{im} T = \dim W.$$

The formula

$$\dim V = \dim \ker T + \dim \operatorname{im} T$$

(the fundamental theorem of linear maps, which is 3.21) 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 of the same dimension. Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_n$ be a basis of W. Let $T \in \operatorname{Hom}(V, W)$ 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 $v_1, \ldots, v_n$ is a basis of V. Also, T is surjective because $w_1, \ldots, w_n$ spans W. Furthermore, $\ker T = \{0\}$ because $w_1, \ldots, w_n$ is linearly independent. Thus T is injective. Because T is injective and surjective, it is an isomorphism (see 3.63). Hence V and W are isomorphic. □

Every finite-dimensional vector space is isomorphic to some $\mathbb{F}^n$. Thus why not just study $\mathbb{F}^n$ instead of more general vector spaces? To answer this question, note that an investigation of $\mathbb{F}^n$ 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 $\mathbb{F}^m$, thinking of them that way often adds complexity but no new insight.

The previous result implies that each finite-dimensional vector space V is isomorphic to $\mathbb{F}^n$, where $n = \dim V$. For example, if m is a nonnegative integer, then $\mathbb{F}[X]_{\le m}$ is isomorphic to $\mathbb{F}^{m+1}$.

Recall that the notation $\mathbb{F}^{m,n}$ denotes the vector space of m-by-n matrices with entries in $\mathbb{F}$. 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 $T \in \text{Hom}(V, W)$, we have a matrix $\text{Mat}(T) \in \mathbb{F}^{m,n}$. Thus once bases have been fixed for V and W, Mat becomes a function from $\text{Hom}(V, W)$ to $\mathbb{F}^{m,n}$. Notice that 3.35 and 3.38 show that Mat is a linear map. This linear map is actually an isomorphism, as we now show.

Result 3.71

$\text{Hom}(V, W)$ and $\mathbb{F}^{m,n}$ 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 Mat is an isomorphism between $\text{Hom}(V, W)$ and $\mathbb{F}^{m,n}$.

Proof. We already noted that Mat is linear. We need to prove that Mat is injective and surjective.

We begin with injectivity. If $T \in \text{Hom}(V, W)$ and $\text{Mat}(T) = 0$, then $Tv_k = 0$ for each $k = 1, \dots, n$. Because $v_1, \dots, v_n$ is a basis of V, this implies $T = 0$. Thus Mat is injective (by 3.15).

To prove that Mat is surjective, suppose $A \in \mathbb{F}^{m,n}$. By the linear map lemma (3.4), there exists $T \in \text{Hom}(V, W)$ such that

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j$$

for each $k = 1, \dots, n$. Because $\text{Mat}(T)$ equals A, the range of Mat equals $\mathbb{F}^{m,n}$, as desired. □

Now we can determine the dimension of the vector space of linear maps from one finite-dimensional vector space to another.

Result 3.72

$\dim \operatorname{Hom}(V, W) = (\dim V)(\dim W)$
Suppose V and W are finite-dimensional. Then $\operatorname{Hom}(V, W)$ is finite-dimensional and

$$\dim \operatorname{Hom}(V, W) = (\dim V)(\dim W).$$

Proof. The desired result follows from 3.71, 3.70, 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.

Definition 3.73

Matrix of a vector, $\operatorname{Mat}(v)$
Suppose $v \in V$ and $v_1, \ldots, v_n$ is a basis of V. The *matrix of* v with respect to this basis is the n-by-1 matrix

$$\operatorname{Mat}(v) = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

where $b_1, \ldots, b_n$ are the scalars such that

$$v = b_1 v_1 + \cdots + b_n v_n.$$

The matrix $\operatorname{Mat}(v)$ of a vector $v \in V$ depends on the basis $v_1, \ldots, v_n$ 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.

Example 3.74 **Matrix of a vector**

- The matrix of the polynomial $2 - 7x + 5x^3 + x^4$ with respect to the standard basis of $\mathbb{R}[X]_{\leq 4}$ is

$$\begin{pmatrix} 2 \\ -7 \\ 0 \\ 5 \\ 1 \end{pmatrix}.$$

- The matrix of a vector $x \in \mathbb{F}^n$ 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

Example 3.74 *continued*

words, if $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$, then

$$\operatorname{Mat}(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Occasionally we want to think of elements of V as relabeled to be n-by-1 matrices. Once a basis $v_1, \ldots, v_n$ is chosen, the function Mat that takes $v \in V$ to $\operatorname{Mat}(v)$ is an isomorphism of V onto $\mathbb{F}^{n,1}$ that implements this relabeling.

Recall that if A is an m-by-n matrix, then $A_{\cdot,k}$ denotes the kth column of A, thought of as an m-by-1 matrix. In the next result, $\operatorname{Mat}(Tv_k)$ is computed with respect to the basis $w_1, \ldots, w_m$ of W.

Result 3.75

$\operatorname{Mat}(T)_{\cdot,k} = \operatorname{Mat}(Tv_k)$

Suppose $T \in \operatorname{Hom}(V, W)$ and $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Let $1 \le k \le n$. Then the kth column of $\operatorname{Mat}(T)$, which is denoted by $\operatorname{Mat}(T)_{\cdot,k}$, equals $\operatorname{Mat}(Tv_k)$.

Proof. The desired result follows immediately from the definitions of $\operatorname{Mat}(T)$ and $\operatorname{Mat}(Tv_k)$. □

The next result shows how the notions of the matrix of a linear map, the matrix of a vector, and matrix multiplication fit together.

Result 3.76

Linear maps act like matrix multiplication

Suppose $T \in \operatorname{Hom}(V, W)$ 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

$$\operatorname{Mat}(Tv) = \operatorname{Mat}(T) \operatorname{Mat}(v).$$

Proof. Suppose $v = b_1 v_1 + \cdots + b_n v_n$, where $b_1, \ldots, b_n \in \mathbb{F}$. Thus

$$Tv = b_1 Tv_1 + \cdots + b_n Tv_n. \tag{3.77}$$

Hence

$$\begin{aligned} \operatorname{Mat}(Tv) &= b_1 \operatorname{Mat}(Tv_1) + \cdots + b_n \operatorname{Mat}(Tv_n) \\ &= b_1 \operatorname{Mat}(T)_{\cdot,1} + \cdots + b_n \operatorname{Mat}(T)_{\cdot,n} \\ &= \operatorname{Mat}(T) \operatorname{Mat}(v), \end{aligned}$$

where the first equality follows from 3.77 and the linearity of Mat, the second equality comes from 3.75, and the last equality comes from 3.50. □

Each m-by-n matrix A induces a linear map from $\mathbb{F}^{n,1}$ to $\mathbb{F}^{m,1}$, namely the matrix multiplication function that takes $x \in \mathbb{F}^{n,1}$ to $Ax \in \mathbb{F}^{m,1}$. The result above can be used

to think of every linear map (from a finite-dimensional vector space to another finite-dimensional vector space) as a matrix multiplication map after suitable relabeling via the isomorphisms given by Mat. Specifically, if $T \in \text{Hom}(V, W)$ and we identify $v \in V$ with $\text{Mat}(v) \in \mathbb{F}^{n,1}$, then the result above says that we can identify Tv with $\text{Mat}(T)\,\text{Mat}(v)$.

Because the result above allows us to think (via isomorphisms) of each linear map as multiplication on $\mathbb{F}^{n,1}$ 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.

Notice that no bases are in sight in the statement of the next result. Although $\text{Mat}(T)$ in the next result depends on a choice of bases of V and W, the next result shows that the column rank of $\text{Mat}(T)$ is the same for all such choices (because $\text{im}\, T$ does not depend on a choice of basis).

Dimension of $\text{im}\, T$ equals column rank of $\text{Mat}(T)$ **Result 3.78**
Suppose V and W are finite-dimensional and $T \in \text{Hom}(V, W)$. Then $\dim \text{im}\, T$ equals the column rank of $\text{Mat}(T)$.

Proof. Suppose $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. The linear map that takes $w \in W$ to $\text{Mat}(w)$ is an isomorphism from W onto the space $\mathbb{F}^{m,1}$ of m-by-1 column vectors. The restriction of this isomorphism to $\text{im}\, T$ [which equals $\text{span}(Tv_1, \dots, Tv_n)$ by Exercise 10 in Section 3B] is an isomorphism from $\text{im}\, T$ onto $\text{span}\big(\text{Mat}(Tv_1), \dots, \text{Mat}(Tv_n)\big)$. For each $k \in \{1, \dots, n\}$, the m-by-1 matrix $\text{Mat}(Tv_k)$ equals column k of $\text{Mat}(T)$. Thus

$$\dim \text{im}\, T = \text{the column rank of } \text{Mat}(T),$$

as desired. □

Change of Basis

In Section 3C we defined the matrix

$$\text{Mat}\big(T, (v_1, \dots, v_n), (w_1, \dots, w_m)\big)$$

of a linear map T from V to a possibly different vector space W, where $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. For linear maps from a vector space to itself, we usually use the same basis for both the domain vector space and the target vector space. When using a single basis in both capacities, we often write the basis only once. In other words, if $T \in \text{End}(V)$ and $v_1, \dots, v_n$ is a basis of V, then the notation $\text{Mat}\big(T, (v_1, \dots, v_n)\big)$ is defined by the equation

$$\text{Mat}\big(T, (v_1, \dots, v_n)\big) = \text{Mat}\big(T, (v_1, \dots, v_n), (v_1, \dots, v_n)\big).$$

If the basis $v_1, \dots, v_n$ is clear from the context, then we can write just $\mathrm{Mat}(T)$.

Identity matrix, I — Definition 3.79

Suppose n is a positive integer. The n-by-n matrix

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

with 1's on the diagonal (the entries where the row number equals the column number) and 0's elsewhere is called the *identity matrix* and is denoted by I.

In the definition above, the 0 in the lower left corner of the matrix indicates that all entries below the diagonal are 0, and the 0 in the upper right corner indicates that all entries above the diagonal are 0.

With respect to each basis of V, the matrix of the identity operator $I \in \mathrm{End}(V)$ is the identity matrix I. Note that the symbol I is used to denote both the identity operator and the identity matrix. The context indicates which meaning of I is intended. For example, consider the equation $\mathrm{Mat}(I) = I$; 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 (meaning it has the same number of rows as columns) with the same size as I, then $AI = IA = A$, as you should verify.

Invertible, inverse, A^{-1} — Definition 3.80

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}.

The same proof as used in 3.60 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).

Some mathematicians use the terms **nonsingular** and **singular**, which mean the same as invertible and noninvertible.

If A is an invertible matrix, then $(A^{-1})^{-1} = A$ because

$$A^{-1}A = AA^{-1} = I.$$

Also, if A and C are invertible square matrices of the same size, then AC is invertible and $(AC)^{-1} = C^{-1}A^{-1}$ because

$$\begin{aligned}(AC)(C^{-1}A^{-1}) &= A(CC^{-1})A^{-1} \\ &= AIA^{-1} \\ &= AA^{-1} \\ &= I,\end{aligned}$$

and similarly $(C^{-1}A^{-1})(AC) = I$.

The next result holds because we defined matrix multiplication to make it true—see 3.43 and the material preceding it. Now we are just being more explicit about the bases involved.

Matrix of product of linear maps **Result 3.81**

Suppose $T \in \text{Hom}(U, V)$ and $S \in \text{Hom}(V, W)$. If $u_1, \dots, u_m$ is a basis of U, $v_1, \dots, v_n$ is a basis of V, and $w_1, \dots, w_p$ is a basis of W, then

$$\text{Mat}\big(ST, (u_1, \dots, u_m), (w_1, \dots, w_p)\big) = \text{Mat}\big(S, (v_1, \dots, v_n), (w_1, \dots, w_p)\big)\,\text{Mat}\big(T, (u_1, \dots, u_m), (v_1, \dots, v_n)\big).$$

The next result deals with the matrix of the identity operator I with respect to two different bases. Note that the kth column of $\text{Mat}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$ consists of the scalars needed to write u_k as a linear combination of the basis $v_1, \dots, v_n$.

In the statement of the next result, I denotes the identity operator from V to V. In the proof, I also denotes the n-by-n identity matrix.

Matrix of identity operator with respect to two bases **Result 3.82**

Suppose that $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then the matrices

$$\text{Mat}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big) \quad \text{and} \quad \text{Mat}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big)$$

are invertible, and each is the inverse of the other.

Proof. In 3.81, replace w_k with u_k, and replace S and T with I, getting

$$I = \text{Mat}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big)\,\text{Mat}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

Now interchange the roles of the u's and v's, getting

$$I = \text{Mat}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)\,\text{Mat}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big).$$

These two equations above give the desired result. □

Example 3.83 **Matrix of identity operator on $\mathbb{F}^2$ with respect to two bases**

Consider the bases $(4, 2), (5, 3)$ and $(1, 0), (0, 1)$ of $\mathbb{F}^2$. Because $I(4, 2) = 4(1, 0) + 2(0, 1)$ and $I(5, 3) = 5(1, 0) + 3(0, 1)$, we have

$$\text{Mat}\Big(I, \big((4, 2), (5, 3)\big), \big((1, 0), (0, 1)\big)\Big) = \begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}.$$

The inverse of the matrix above is

$$\begin{pmatrix} 3/2 & -5/2 \\ -1 & 2 \end{pmatrix},$$

as you should verify. Thus 3.82 implies that

$$\text{Mat}\Big(I, \big((1, 0), (0, 1)\big), \big((4, 2), (5, 3)\big)\Big) = \begin{pmatrix} 3/2 & -5/2 \\ -1 & 2 \end{pmatrix}.$$

Our next result shows how the matrix of T changes when we change bases. In the next result, we have two different bases of V, each of which is used as a basis for the domain space and as a basis for the target space. Recall our shorthand notation that allows us to display a basis only once when it is used in both capacities:

$$\text{Mat}\big(T, (u_1, \dots, u_n)\big) = \text{Mat}\big(T, (u_1, \dots, u_n), (u_1, \dots, u_n)\big).$$

Change-of-basis formula **Result 3.84**

Suppose $T \in \text{End}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Let

$$A = \text{Mat}\big(T, (u_1, \dots, u_n)\big) \quad \text{and} \quad B = \text{Mat}\big(T, (v_1, \dots, v_n)\big)$$

and $C = \text{Mat}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$. Then

$$A = C^{-1}BC.$$

Proof. In 3.81, replace w_k with u_k and replace S with I, getting

$$A = C^{-1}\,\text{Mat}\big(T, (u_1, \dots, u_n), (v_1, \dots, v_n)\big), \tag{3.85}$$

where we have used 3.82.

Again use 3.81, this time replacing w_k with v_k. Also replace T with I and replace S with T, getting

$$\text{Mat}\big(T, (u_1, \dots, u_n), (v_1, \dots, v_n)\big) = BC.$$

Substituting the equation above into 3.85 gives the equation $A = C^{-1}BC$. □

The proof of the next result is left as an exercise.

Matrix of inverse equals inverse of matrix **Result 3.86**

Suppose that $v_1, \dots, v_n$ is a basis of V and $T \in \text{End}(V)$ is invertible. Then $\text{Mat}(T^{-1}) = \big(\text{Mat}(T)\big)^{-1}$, where both matrices are with respect to the basis $v_1, \dots, v_n$.

Exercises

3D.1 Suppose $T \in \text{Hom}(V, W)$ is invertible. Show that T^{-1} is invertible and

$$(T^{-1})^{-1} = T.$$

3D.2 Suppose $T \in \text{Hom}(U, V)$ and $S \in \text{Hom}(V, W)$ are both invertible linear maps. Prove that $ST \in \text{Hom}(U, W)$ is invertible and that $(ST)^{-1} = T^{-1}S^{-1}$.

3D.3 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that the following are equivalent.

(a) T is invertible.

(b) $Tv_1, \ldots, Tv_n$ is a basis of V for every basis $v_1, \ldots, v_n$ of V.

(c) $Tv_1, \ldots, Tv_n$ is a basis of V for some basis $v_1, \ldots, v_n$ of V.

3D.4 Suppose V is finite-dimensional and $\dim V > 1$. Prove that the set of noninvertible linear maps from V to itself is not a subspace of $\text{End}(V)$.

3D.5 Suppose V is finite-dimensional, U is a subspace of V, and $S \in \text{Hom}(U, V)$. Prove that there exists an invertible linear map T from V to itself such that $Tu = Su$ for every $u \in U$ if and only if S is injective.

3D.6 Suppose that W is finite-dimensional and $S, T \in \text{Hom}(V, W)$. Prove that $\ker S = \ker T$ if and only if there exists an invertible $E \in \text{End}(W)$ such that $S = ET$.

3D.7 Suppose that V is finite-dimensional and $S, T \in \text{Hom}(V, W)$. Prove that $\text{im}\, S = \text{im}\, T$ if and only if there exists an invertible $E \in \text{End}(V)$ such that $S = TE$.

3D.8 Suppose V and W are finite-dimensional and $S, T \in \text{Hom}(V, W)$. Prove that there exist invertible $E_1 \in \text{End}(V)$ and $E_2 \in \text{End}(W)$ such that $S = E_2 T E_1$ if and only if $\dim \ker S = \dim \ker T$.

3D.9 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. Thus $T|_U$ is the function whose domain is U, with $T|_U$ defined by $T|_U(u) = Tu$ for every $u \in U$.

3D.10 Suppose V and W are finite-dimensional and U is a subspace of V. Let

$$\mathscr{E} = \{T \in \text{Hom}(V, W) : U \subseteq \ker T\}.$$

(a) Show that $\mathscr{E}$ is a subspace of $\text{Hom}(V, W)$.

(b) Find a formula for $\dim \mathscr{E}$ in terms of $\dim V$, $\dim W$, and $\dim U$.

Hint: Define $\Phi: \text{Hom}(V, W) \to \text{Hom}(U, W)$ by $\Phi(T) = T|_U$. What is $\ker \Phi$? What is $\text{im}\, \Phi$?

3D.11 Suppose V is finite-dimensional and $S, T \in \text{End}(V)$. Prove that

$$ST \text{ is invertible} \iff S \text{ and } T \text{ are invertible.}$$

3D.12 Suppose V is finite-dimensional and $S, T, U \in \operatorname{End}(V)$ and $STU = I$. Show that T is invertible and that $T^{-1} = US$.

3D.13 Show that the result in Exercise 12 can fail without the hypothesis that V is finite-dimensional.

3D.14 Prove or give a counterexample: If V is a finite-dimensional vector space and $R, S, T \in \operatorname{End}(V)$ are such that RST is surjective, then S is injective.

3D.15 Suppose $T \in \operatorname{End}(V)$ and $v_1, \dots, v_m$ is a list in V such that $Tv_1, \dots, Tv_m$ spans V. Prove that $v_1, \dots, v_m$ spans V.

3D.16 Prove that every linear map from $\mathbb{F}^{n,1}$ to $\mathbb{F}^{m,1}$ is given by a matrix multiplication. In other words, prove that if $T \in \operatorname{Hom}(\mathbb{F}^{n,1}, \mathbb{F}^{m,1})$, then there exists an m-by-n matrix A such that $Tx = Ax$ for every $x \in \mathbb{F}^{n,1}$.

3D.17 Suppose V is finite-dimensional and $S \in \operatorname{End}(V)$. Define $\mathcal{A} \in \operatorname{End}\big(\operatorname{End}(V)\big)$ by

$$\mathcal{A}(T) = ST$$

for $T \in \operatorname{End}(V)$.

(a) Show that $\dim \ker \mathcal{A} = (\dim V)(\dim \ker S)$.

(b) Show that $\dim \operatorname{im} \mathcal{A} = (\dim V)(\dim \operatorname{im} S)$.

3D.18 Show that V and $\operatorname{Hom}(\mathbb{F}, V)$ are isomorphic vector spaces.

3D.19 Suppose V is finite-dimensional and $T \in \operatorname{End}(V)$. Prove that T has the same matrix with respect to every basis of V if and only if T is a scalar multiple of the identity operator.

3D.20 Suppose $q \in \mathbb{R}[X]$. Prove that there exists a polynomial $p \in \mathbb{R}[X]$ such that

$$q(x) = (x^2 + x)p''(x) + 2xp'(x) + p(3)$$

for all $x \in \mathbb{R}$.

3D.21 Suppose n is a positive integer and $A_{j,k} \in \mathbb{F}$ for all $j, k = 1, \dots, n$. 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

$$\sum_{k=1}^{n} A_{1,k} x_k = 0$$
$$\vdots$$
$$\sum_{k=1}^{n} A_{n,k} x_k = 0.$$

(b) For every $c_1, \ldots, c_n \in \mathbb{F}$, there exists a solution to the system of equations

$$\sum_{k=1}^{n} A_{1,k} x_k = c_1$$
$$\vdots$$
$$\sum_{k=1}^{n} A_{n,k} x_k = c_n.$$

3D.22 Suppose $T \in \text{End}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Prove that

$$\text{Mat}\big(T, (v_1, \ldots, v_n)\big) \text{ is invertible} \iff T \text{ is invertible.}$$

3D.23 Suppose that $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Let $T \in \text{End}(V)$ be such that $Tv_k = u_k$ for each $k = 1, \ldots, n$. Prove that

$$\text{Mat}\big(T, (v_1, \ldots, v_n)\big) = \text{Mat}\big(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)\big).$$

3D.24 Suppose A and B are square matrices of the same size and $AB = I$. Prove that $BA = I$.

3E Products and Quotients of Vector Spaces

Products of Vector Spaces

As usual when dealing with more than one vector space, all vector spaces in use should be over the same field.

Product of vector spaces — Definition 3.87

Suppose $V_1, \dots, V_m$ are vector spaces over $\mathbb{F}$.

- The *product* $V_1 \times \cdots \times V_m$ is defined by

$$V_1 \times \cdots \times V_m = \{(v_1, \dots, v_m) : v_1 \in V_1, \dots, v_m \in V_m\}.$$

- Addition on $V_1 \times \cdots \times V_m$ is defined by

$$(u_1, \dots, u_m) + (v_1, \dots, v_m) = (u_1 + v_1, \dots, u_m + v_m).$$

- Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by

$$\lambda(v_1, \dots, v_m) = (\lambda v_1, \dots, \lambda v_m).$$

Example 3.88 **Product of the vector spaces $\mathbb{R}[X]_{\le 5}$ and $\mathbb{R}^3$**

Elements of $\mathbb{R}[X]_{\le 5} \times \mathbb{R}^3$ are lists of length two, with the first item in the list an element of $\mathbb{R}[X]_{\le 5}$ and the second item in the list an element of $\mathbb{R}^3$.

For example, $\big(5 - 6x + 4x^2, (3, 8, 7)\big)$ and $\big(x + 9x^5, (2, 2, 2)\big)$ are elements of $\mathbb{R}[X]_{\le 5} \times \mathbb{R}^3$. Their sum is defined by

$$\begin{aligned}&\big(5 - 6x + 4x^2, (3, 8, 7)\big) + \big(x + 9x^5, (2, 2, 2)\big)\\ &\quad = \big(5 - 5x + 4x^2 + 9x^5, (5, 10, 9)\big).\end{aligned}$$

Also, $2\big(5 - 6x + 4x^2, (3, 8, 7)\big) = \big(10 - 12x + 8x^2, (6, 16, 14)\big)$.

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 by 3.87.

Product of vector spaces is a vector space — Result 3.89

Suppose $V_1, \dots, V_m$ are vector spaces over $\mathbb{F}$. Then $V_1 \times \cdots \times V_m$ is a vector space over $\mathbb{F}$.

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 $(0, \dots, 0)$, where the 0 in the kth slot is the additive identity of V_k. The additive inverse of $(v_1, \dots, v_m) \in V_1 \times \cdots \times V_m$ is $(-v_1, \dots, -v_m)$.

Example 3.90 **$\mathbb{R}^2 \times \mathbb{R}^3 \ne \mathbb{R}^5$ but $\mathbb{R}^2 \times \mathbb{R}^3$ is isomorphic to $\mathbb{R}^5$**

Elements of the vector space $\mathbb{R}^2 \times \mathbb{R}^3$ are lists

$$\big((x_1, x_2), (x_3, x_4, x_5)\big),$$

Example 3.90 *continued*

where $x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}$. Elements of $\mathbb{R}^5$ are lists

$$(x_1, x_2, x_3, x_4, x_5),$$

where $x_1, x_2, x_3, x_4, x_5 \in \mathbb{R}$.

Although elements of $\mathbb{R}^2 \times \mathbb{R}^3$ and $\mathbb{R}^5$ look similar, they are not the same kind of object. Elements of $\mathbb{R}^2 \times \mathbb{R}^3$ are lists of length two (with the first item itself a list of length two and the second item a list of length three), and elements of $\mathbb{R}^5$ are lists of length five. Thus $\mathbb{R}^2 \times \mathbb{R}^3$ does not equal $\mathbb{R}^5$.

The linear map

$$\big((x_1, x_2), (x_3, x_4, x_5)\big) \mapsto (x_1, x_2, x_3, x_4, x_5)$$

is an isomorphism of the vector space $\mathbb{R}^2 \times \mathbb{R}^3$ onto the vector space $\mathbb{R}^5$. Thus these two vector spaces are isomorphic, although they are not equal.

This isomorphism is so natural that we should think of it as a relabeling. Some people informally say that $\mathbb{R}^2 \times \mathbb{R}^3$ equals $\mathbb{R}^5$, which is not technically correct but which captures the spirit of identification via relabeling.

The next example illustrates the idea that we will use in the proof of 3.92.

Example 3.91 **A basis of $\mathbb{R}[X]_{\le 2} \times \mathbb{R}^2$**

Consider this list of length five of elements of $\mathbb{R}[X]_{\le 2} \times \mathbb{R}^2$:

$$\big(1, (0,0)\big),\ \big(x, (0,0)\big),\ \big(x^2, (0,0)\big),\ \big(0, (1,0)\big),\ \big(0, (0,1)\big).$$

The list above is linearly independent and it spans $\mathbb{R}[X]_{\le 2} \times \mathbb{R}^2$. Thus it is a basis of $\mathbb{R}[X]_{\le 2} \times \mathbb{R}^2$.

Result 3.92

Dimension of a product is the sum of dimensions
Suppose $V_1, \dots, V_m$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m.$$

Proof. Choose a basis of each V_k. For each basis vector of each V_k, consider the element of $V_1 \times \cdots \times V_m$ that equals the basis vector in the kth slot and 0 in the other slots. The list of all such vectors is linearly independent and spans $V_1 \times \cdots \times V_m$. Thus it is a basis of $V_1 \times \cdots \times V_m$. The length of this basis is $\dim V_1 + \cdots + \dim V_m$, as desired. □

In the next result, the map Γ is surjective by the definition of $V_1 + \cdots + V_m$. Thus the last word in the result below could be changed from "injective" to "invertible".

Products and direct sums — Result 3.93

Suppose that $V_1, \dots, V_m$ are subspaces of V. Define a linear map $\Gamma: V_1 \times \cdots \times V_m \to V_1 + \cdots + V_m$ by

$$\Gamma(v_1, \dots, v_m) = v_1 + \cdots + v_m.$$

Then $V_1 + \cdots + V_m$ is a direct sum if and only if Γ is injective.

Proof. By 3.15, Γ is injective if and only if the only way to write 0 as a sum $v_1 + \cdots + v_m$, where each v_k is in V_k, is by taking each v_k equal to 0. Thus 1.45 shows that Γ is injective if and only if $V_1 + \cdots + V_m$ is a direct sum, as desired. □

A sum is a direct sum if and only if dimensions add up — Result 3.94

Suppose V is finite-dimensional and $V_1, \dots, V_m$ are subspaces of V. Then $V_1 + \cdots + V_m$ is a direct sum if and only if

$$\dim(V_1 + \cdots + V_m) = \dim V_1 + \cdots + \dim V_m.$$

Proof. The map Γ in 3.93 is surjective. Thus by the fundamental theorem of linear maps (3.21), Γ is injective if and only if

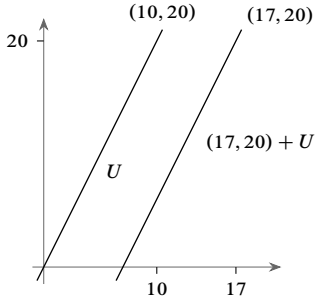
$$\dim(V_1 + \cdots + V_m) = \dim(V_1 \times \cdots \times V_m).$$

Combining 3.93 and 3.92 now shows that $V_1 + \cdots + V_m$ is a direct sum if and only if

$$\dim(V_1 + \cdots + V_m) = \dim V_1 + \cdots + \dim V_m,$$

as desired. □

In the special case $m = 2$, an alternative proof that $V_1 + V_2$ is a direct sum if and only if $\dim(V_1 + V_2) = \dim V_1 + \dim V_2$ can be obtained by combining 1.46 and 2.43.



$(17, 20) + U$ *is parallel to the subspace* U.

Quotient Spaces

We begin our approach to quotient spaces by defining the sum of a vector and a subset.

$v + U$ — Notation 3.95

Suppose $v \in V$ and $U \subseteq V$. Then $v + U$ is the subset of V defined by

$$v + U = \{v + u : u \in U\}.$$

Example 3.96 **Sum of a vector and a one-dimensional subspace of $\mathbb{R}^2$**

Example 3.96 *continued*

Suppose

$$U = \{(x, 2x) \in \mathbb{R}^2 : x \in \mathbb{R}\}.$$

Hence U is the line in $\mathbb{R}^2$ through the origin with slope 2. Thus

$$(17, 20) + U$$

is the line in $\mathbb{R}^2$ that contains the point $(17, 20)$ and has slope 2.

Because

$$(10, 20) \in U \quad \text{and} \quad (17, 20) \in (17, 20) + U,$$

we see that $(17, 20) + U$ is obtained by moving U to the right by 7 units.

Translate Definition 3.97

For $v \in V$ and U a subset of V, the set $v + U$ is said to be a *translate* of U.

Example 3.98 Translates

- If U is the line in $\mathbb{R}^2$ defined by $U = \{(x, 2x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$, then all lines in $\mathbb{R}^2$ with slope 2 are translates of U. See Example 3.96 above for a drawing of U and one of its translates.
- More generally, if U is a line in $\mathbb{R}^2$, then the set of all translates of U is the set of all lines in $\mathbb{R}^2$ that are parallel to U.
- If $U = \{(x, y, 0) \in \mathbb{R}^3 : x, y \in \mathbb{R}\}$, then the translates of U are the planes in $\mathbb{R}^3$ that are parallel to the xy-plane U.
- More generally, if U is a plane in $\mathbb{R}^3$, then the set of all translates of U is the set of all planes in $\mathbb{R}^3$ that are parallel to U (see, for example, Exercise 7).

Quotient space, V/U Definition 3.99

Suppose U is a subspace of V. Then the *quotient space* V/U is the set of all translates of U. Thus

$$V/U = \{v + U : v \in V\}.$$

Example 3.100 Quotient spaces

Example 3.100
continued

- If $U = \{(x, 2x) \in \mathbb{R}^2 : x \in \mathbb{R}\}$, then $\mathbb{R}^2/U$ is the set of all lines in $\mathbb{R}^2$ that have slope 2.
- If U is a line in $\mathbb{R}^3$ containing the origin, then $\mathbb{R}^3/U$ is the set of all lines in $\mathbb{R}^3$ parallel to U.
- If U is a plane in $\mathbb{R}^3$ containing the origin, then $\mathbb{R}^3/U$ is the set of all planes in $\mathbb{R}^3$ parallel to U.

Our next goal is to make V/U into a vector space. To do this, we will need the next result.

Two translates of a subspace are equal or disjoint Result 3.101

Suppose U is a subspace of V and $v, w \in V$. Then

$$v - w \in U \iff v + U = w + U \iff (v + U) \cap (w + U) \neq \varnothing.$$

Proof. First suppose $v - w \in U$. If $u \in U$, then

$$v + u = w + \big((v - w) + u\big) \in w + U.$$

Thus $v + U \subseteq w + U$. Similarly, $w + U \subseteq v + U$. Thus $v + U = w + U$, completing the proof that $v - w \in U$ implies $v + U = w + U$.

The equation $v + U = w + U$ implies that $(v + U) \cap (w + U) \neq \varnothing$.

Now suppose $(v + U) \cap (w + U) \neq \varnothing$. 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 $(v + U) \cap (w + U) \neq \varnothing$ implies $v - w \in U$, which completes the proof. □

Now we can define addition and scalar multiplication on V/U.

Addition and scalar multiplication on V/U Definition 3.102

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 all $v, w \in V$ and all $\lambda \in \mathbb{F}$.

As part of the proof of the next result, we will show that the definitions above make sense.

Quotient space is a vector space — Result 3.103

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 a translate of U is not unique. Specifically, suppose $v_1, v_2, w_1, w_2 \in V$ are such that

$$v_1 + U = v_2 + U \quad \text{and} \quad w_1 + U = w_2 + U.$$

To show that the definition of addition on V/U given above makes sense, we must show that $(v_1 + w_1) + U = (v_2 + w_2) + U$.

By 3.101, we have

$$v_1 - v_2 \in U \quad \text{and} \quad w_1 - w_2 \in U.$$

Because U is a subspace of V and thus is closed under addition, this implies that $(v_1 - v_2) + (w_1 - w_2) \in U$. Thus $(v_1 + w_1) - (v_2 + w_2) \in U$. Using 3.101 again, we see that

$$(v_1 + w_1) + U = (v_2 + w_2) + U,$$

as desired. Thus the definition of addition on V/U makes sense.

Similarly, suppose $\lambda \in \mathbb{F}$. We are still assuming that $v_1 + U = v_2 + U$. Because U is a subspace of V and thus is closed under scalar multiplication, we have $\lambda(v_1 - v_2) \in U$. Thus $\lambda v_1 - \lambda v_2 \in U$. Hence 3.101 implies that

$$(\lambda v_1) + U = (\lambda v_2) + U.$$

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 lead to a computation of the dimension of V/U.

Quotient map, π — Definition 3.104

Suppose U is a subspace of V. The *quotient map* $\pi: V \to V/U$ is the linear map defined by

$$\pi(v) = v + U$$

for each $v \in V$.

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.

Dimension of quotient space — Result 3.105

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim V/U = \dim V - \dim U.$$

Proof. Let π denote the quotient map from V to V/U. If $v \in V$, then $v + U = 0 + U$ if and only if $v \in U$ (by 3.101), which implies that $\ker \pi = U$. The definition of π implies $\operatorname{im} \pi = V/U$. The fundamental theorem of linear maps (3.21) now implies $\dim V = \dim U + \dim V/U$, which gives the desired result. □

Each linear map T on V induces a linear map $\widetilde{T}$ on $V/(\ker T)$, as defined below.

$\widetilde{T}$ — Notation 3.106

Suppose $T \in \operatorname{Hom}(V, W)$. Define $\widetilde{T}: V/(\ker T) \to W$ by

$$\widetilde{T}(v + \ker T) = Tv.$$

To show that the definition of $\widetilde{T}$ makes sense, suppose $u, v \in V$ are such that $u + \ker T = v + \ker T$. By 3.101, we have $u - v \in \ker T$. Thus $T(u - v) = 0$. Hence $Tu = Tv$. Thus the definition of $\widetilde{T}$ indeed makes sense. The routine verification that $\widetilde{T}$ is a linear map from $V/(\ker T)$ to W is left to the reader.

The next result shows that we can think of $\widetilde{T}$ as a modified version of T, with a domain that produces a one-to-one map.

Null space and range of $\widetilde{T}$ — Result 3.107

Suppose $T \in \operatorname{Hom}(V, W)$. Then

(a) $\widetilde{T} \circ \pi = T$, where π is the quotient map of V onto $V/(\ker T)$;

(b) $\widetilde{T}$ is injective;

(c) $\operatorname{im} \widetilde{T} = \operatorname{im} T$;

(d) $V/(\ker T)$ and $\operatorname{im} T$ are isomorphic vector spaces.

Proof.

(a) If $v \in V$, then $(\widetilde{T} \circ \pi)(v) = \widetilde{T}\big(\pi(v)\big) = \widetilde{T}(v + \ker T) = Tv$, as desired.

(b) Suppose $v \in V$ and $\widetilde{T}(v + \ker T) = 0$. Then $Tv = 0$. Thus $v \in \ker T$. Hence 3.101 implies that $v + \ker T = 0 + \ker T$. This implies that $\ker \widetilde{T} = \{0 + \ker T\}$. Hence $\widetilde{T}$ is injective, as desired.

(c) The definition of $\widetilde{T}$ shows that $\operatorname{im} \widetilde{T} = \operatorname{im} T$.

(d) Now (b) and (c) imply that if we think of $\widetilde{T}$ as mapping into $\operatorname{im} T$, then $\widetilde{T}$ is an isomorphism from $V/(\ker T)$ onto $\operatorname{im} T$. □

Exercises

3E.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 this exercise 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 $V \times W$.

3E.2 Suppose that $V_1, \dots, V_m$ are vector spaces such that $V_1 \times \cdots \times V_m$ is finite-dimensional. Prove that V_k is finite-dimensional for each $k = 1, \dots, m$.

3E.3 Suppose $V_1, \dots, V_m$ are vector spaces. Prove that $\text{Hom}(V_1 \times \cdots \times V_m, W)$ and $\text{Hom}(V_1, W) \times \cdots \times \text{Hom}(V_m, W)$ are isomorphic vector spaces.

There is no assumption in the exercise above or in the two following exercises that the vector spaces are finite-dimensional.

3E.4 Suppose $W_1, \dots, W_m$ are vector spaces. Prove that $\text{Hom}(V, W_1 \times \cdots \times W_m)$ and $\text{Hom}(V, W_1) \times \cdots \times \text{Hom}(V, W_m)$ are isomorphic vector spaces.

3E.5 For m a positive integer, define V^m by

$$V^m = \underbrace{V \times \cdots \times V}_{m \text{ times}}.$$

Prove that V^m and $\text{Hom}(\mathbb{F}^m, V)$ are isomorphic vector spaces.

3E.6 Suppose that v, x are vectors in V and that U, W are subspaces of V such that $v + U = x + W$. Prove that $U = W$.

3E.7 Let $U = \{(x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = 0\}$. Suppose $A \subseteq \mathbb{R}^3$. Prove that A is a translate of U if and only if there exists $c \in \mathbb{R}$ such that

$$A = \{(x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = c\}.$$

3E.8

(a) Suppose $T \in \text{Hom}(V, W)$ and $c \in W$. Prove that $\{x \in V : Tx = c\}$ is either the empty set or is a translate of $\ker T$.

(b) Explain why the set of solutions to a system of linear equations such as 3.27 is either the empty set or is a translate of some subspace of $\mathbb{F}^n$.

3E.9 Prove that a nonempty subset A of V is a translate of some subspace of V if and only if $\lambda v + (1-\lambda)w \in A$ for all $v, w \in A$ and all $\lambda \in \mathbb{F}$.

3E.10 Suppose $A_1 = v + U_1$ and $A_2 = w + U_2$ for some $v, w \in V$ and some subspaces U_1, U_2 of V. Prove that the intersection $A_1 \cap A_2$ is either a translate of some subspace of V or is the empty set.

3E.11 Suppose $U = \{(x_1, x_2, \dots) \in \mathbb{F}^\infty : x_k \neq 0 \text{ for only finitely many } k\}$.

(a) Show that U is a subspace of $\mathbb{F}^\infty$.

(b) Prove that $\mathbb{F}^\infty / U$ is infinite-dimensional.

3E.12 Suppose $v_1, \dots, v_m \in V$. Let

$$A = \{\lambda_1 v_1 + \cdots + \lambda_m v_m : \lambda_1, \dots, \lambda_m \in \mathbb{F} \text{ and } \lambda_1 + \cdots + \lambda_m = 1\}.$$

(a) Prove that A is a translate of some subspace of V.

(b) Prove that if B is a translate of some subspace of V and $\{v_1, \dots, v_m\} \subseteq B$, then $A \subseteq B$.

(c) Prove that A is a translate of some subspace of V of dimension less than m.

3E.13 Suppose U is a subspace of V such that V/U is finite-dimensional. Prove that V is isomorphic to $U \times (V/U)$.

3E.14 Suppose U and W are subspaces of V and $V = U \oplus W$. Suppose $w_1, \dots, w_m$ is a basis of W. Prove that $w_1 + U, \dots, w_m + U$ is a basis of V/U.

3E.15 Suppose U is a subspace of V and $v_1 + U, \dots, v_m + U$ is a basis of V/U and $u_1, \dots, u_n$ is a basis of U. Prove that $v_1, \dots, v_m, u_1, \dots, u_n$ is a basis of V.

3E.16 Suppose $\varphi \in \text{Hom}(V, \mathbb{F})$ and $\varphi \neq 0$. Prove that $\dim V/(\ker \varphi) = 1$.

3E.17 Suppose U is a subspace of V such that $\dim V/U = 1$. Prove that there exists $\varphi \in \text{Hom}(V, \mathbb{F})$ such that $\ker \varphi = U$.

3E.18 Suppose that U is a subspace of V such that V/U is finite-dimensional.

(a) Show that if W is a finite-dimensional subspace of V and $V = U + W$, then $\dim W \geq \dim V/U$.

(b) Prove that there exists a finite-dimensional subspace W of V such that $\dim W = \dim V/U$ and $V = U \oplus W$.

3E.19 Suppose $T \in \text{Hom}(V, W)$ and U is a subspace of V. Let π denote the quotient map from V onto V/U. Prove that there exists $S \in \text{Hom}(V/U, W)$ such that $T = S \circ \pi$ if and only if $U \subseteq \ker T$.

3F Duality

Dual Space and Dual Map

Linear maps into the scalar field $\mathbb{F}$ play a special role in linear algebra, and thus they get a special name.

Linear functional — Definition 3.108

A *linear functional* on V is a linear map from V to $\mathbb{F}$. In other words, a linear functional is an element of $\text{Hom}(V, \mathbb{F})$.

Example 3.109 **Linear functionals**

- Define $\varphi: \mathbb{R}^3 \to \mathbb{R}$ by $\varphi(x, y, z) = 4x - 5y + 2z$. Then φ is a linear functional on $\mathbb{R}^3$.
- Fix $(c_1, \ldots, c_n) \in \mathbb{F}^n$. Define $\varphi: \mathbb{F}^n \to \mathbb{F}$ by $\varphi(x_1, \ldots, x_n) = c_1 x_1 + \cdots + c_n x_n$. Then φ is a linear functional on $\mathbb{F}^n$.
- Define $\varphi: \mathbb{R}[X] \to \mathbb{R}$ by
$$\varphi(p) = 3p''(5) + 7p(4).$$
Then φ is a linear functional on $\mathbb{R}[X]$.
- Define $\varphi: \mathbb{R}[X] \to \mathbb{R}$ by
$$\varphi(p) = \int_0^1 p$$
for each $p \in \mathbb{R}[X]$. Then φ is a linear functional on $\mathbb{R}[X]$.

The vector space $\text{Hom}(V, \mathbb{F})$ also gets a special name and special notation.

Dual space, V' — Definition 3.110

The *dual space* of V, denoted by V', is the vector space of all linear functionals on V. In other words, $V' = \text{Hom}(V, \mathbb{F})$.

Result 3.111

$\dim V' = \dim V$

Suppose V is finite-dimensional. Then V' is also finite-dimensional and

$$\dim V' = \dim V.$$

Proof. By 3.72 we have

$$\dim V' = \dim \operatorname{Hom}(V, \mathbb{F}) = (\dim V)(\dim \mathbb{F}) = \dim V,$$

as desired. □

In the following definition, the linear map lemma (3.4) implies that each φ_j is well defined.

Definition 3.112

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

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Example 3.113 **The dual basis of the standard basis of $\mathbb{F}^n$**

Suppose n is a positive integer. For $1 \leq j \leq n$, define φ_j to be the linear functional on $\mathbb{F}^n$ that selects the jth coordinate of a vector in $\mathbb{F}^n$. Thus

$$\varphi_j(x_1, \dots, x_n) = x_j$$

for each $(x_1, \dots, x_n) \in \mathbb{F}^n$.

Let $e_1, \dots, e_n$ be the standard basis of $\mathbb{F}^n$. Then

$$\varphi_j(e_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

Thus $\varphi_1, \dots, \varphi_n$ is the dual basis of the standard basis $e_1, \dots, e_n$ of $\mathbb{F}^n$.

The next result shows that the dual basis of a basis of V consists of the linear functionals on V that give the coefficients for expressing a vector in V as a linear combination of the basis vectors.

Result 3.114

Dual basis gives coefficients for linear combination

Suppose $v_1, \dots, v_n$ is a basis of V and $\varphi_1, \dots, \varphi_n$ is the dual basis. Then

$$v = \varphi_1(v)v_1 + \cdots + \varphi_n(v)v_n$$

for each $v \in V$.

Proof. Suppose $v \in V$. Then there exist $c_1, \dots, c_n \in \mathbb{F}$ such that

$$v = c_1 v_1 + \cdots + c_n v_n. \tag{3.115}$$

If $j \in \{1, \dots, n\}$, then applying φ_j to both sides of the equation above gives

$$\varphi_j(v) = c_j.$$

Substituting the values for $c_1, \dots, c_n$ given by the equation above into 3.115 shows that $v = \varphi_1(v)v_1 + \cdots + \varphi_n(v)v_n$. □

The next result shows that the dual basis is indeed a basis of the dual space. Thus the terminology "dual basis" is justified.

Dual basis is a basis of the dual space **Result 3.116**

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 $a_1, \dots, a_n \in \mathbb{F}$ are such that

$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0. \tag{3.117}$$

Now

$$(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_k) = a_k$$

for each $k = 1, \dots, n$. Thus 3.117 shows that $a_1 = \cdots = a_n = 0$. Hence $\varphi_1, \dots, \varphi_n$ is linearly independent.

Because $\varphi_1, \dots, \varphi_n$ is a linearly independent list in V' whose length equals $\dim V'$ (by 3.111), we can conclude that $\varphi_1, \dots, \varphi_n$ is a basis of V' (see 2.38). □

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'.

Dual map, T' **Definition 3.118**

Suppose $T \in \text{Hom}(V, W)$. The *dual map* of T is the linear map $T' \in \text{Hom}(W', V')$ defined for each $\varphi \in W'$ by

$$T'(\varphi) = \varphi \circ T.$$

If $T \in \text{Hom}(V, W)$ 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 $\mathbb{F}$; in other words, $T'(\varphi) \in V'$.

The following two bullet points show that T' is a linear map from W' to V'.

- If $\varphi, \psi \in W'$, then

$$T'(\varphi + \psi) = (\varphi + \psi) \circ T = \varphi \circ T + \psi \circ T = T'(\varphi) + T'(\psi).$$

- If $\lambda \in \mathbb{F}$ and $\varphi \in W'$, then

$$T'(\lambda\varphi) = (\lambda\varphi) \circ T = \lambda(\varphi \circ T) = \lambda T'(\varphi).$$

The prime notation appears with two unrelated meanings in the next example: D' denotes the dual of the linear map D, and p' denotes the derivative of a polynomial p.

Example 3.119 **Dual map of the differentiation linear map**

Define $D: \mathbb{R}[X] \to \mathbb{R}[X]$ by $Dp = p'$.

- Suppose φ is the linear functional on $\mathbb{R}[X]$ defined by $\varphi(p) = p(3)$. Then $D'(\varphi)$ is the linear functional on $\mathbb{R}[X]$ given by

$$\big(D'(\varphi)\big)(p) = (\varphi \circ D)(p) = \varphi(Dp) = \varphi(p') = p'(3).$$

  Thus $D'(\varphi)$ is the linear functional on $\mathbb{R}[X]$ taking p to $p'(3)$.

- Suppose φ is the linear functional on $\mathbb{R}[X]$ defined by $\varphi(p) = \int_0^1 p$. Then $D'(\varphi)$ is the linear functional on $\mathbb{R}[X]$ given by

$$\begin{aligned}\big(D'(\varphi)\big)(p) &= (\varphi \circ D)(p)\\ &= \varphi(Dp)\\ &= \varphi(p')\\ &= \int_0^1 p'\\ &= p(1) - p(0).\end{aligned}$$

  Thus $D'(\varphi)$ is the linear functional on $\mathbb{R}[X]$ taking p to $p(1) - p(0)$.

In the next result, (a) and (b) imply that the function that takes T to T' is a linear map from $\mathrm{Hom}(V, W)$ to $\mathrm{Hom}(W', V')$.

In (c) below, note the reversal of order from ST on the left to $T'S'$ on the right.

Result 3.120 **Algebraic properties of dual maps**

Suppose $T \in \mathrm{Hom}(V, W)$. Then

(a) $(S + T)' = S' + T'$ for all $S \in \mathrm{Hom}(V, W)$;

(b) $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbb{F}$;

(c) $(ST)' = T'S'$ for all $S \in \mathrm{Hom}(W, U)$.

Proof. The proofs of (a) and (b) are left to the reader.

To prove (c), suppose $\varphi \in U'$. Then

$$(ST)'(\varphi) = \varphi \circ (ST) = (\varphi \circ S) \circ T = T'(\varphi \circ S) = T'\big(S'(\varphi)\big) = (T'S')(\varphi),$$

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.

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 equation above shows that $(ST)'(\varphi) = (T'S')(\varphi)$ for all $\varphi \in U'$. Thus $(ST)' = T'S'$. □

Null Space and Range of Dual of Linear Map

Our goal in this subsection is to describe $\ker T'$ and $\operatorname{im} T'$ in terms of $\operatorname{im} T$ and $\ker T$. To do this, we will need the next definition.

Definition 3.121

Annihilator, U^0

For $U \subseteq V$, the *annihilator* of U, denoted by U^0, is defined by

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}.$$

Example 3.122 **Element of an annihilator**

Suppose U is the subspace of $\mathbb{R}[X]$ consisting of polynomial multiples of x^2. If φ is the linear functional on $\mathbb{R}[X]$ defined by $\varphi(p) = p'(0)$, then $\varphi \in U^0$.

For $U \subseteq V$, 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.

Example 3.123 **The annihilator of a two-dimensional subspace of $\mathbb{R}^5$**

Let e_1, e_2, e_3, e_4, e_5 denote the standard basis of $\mathbb{R}^5$; let $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5 \in (\mathbb{R}^5)'$ denote the dual basis of e_1, e_2, e_3, e_4, e_5. Suppose

$$U = \operatorname{span}(e_1, e_2) = \{(x_1, x_2, 0, 0, 0) \in \mathbb{R}^5 : x_1, x_2 \in \mathbb{R}\}.$$

We want to show that $U^0 = \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$.

Recall (see 3.113) that φ_j is the linear functional on $\mathbb{R}^5$ that selects the jth 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 $c_3, c_4, c_5 \in \mathbb{R}$ 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 $\varphi \in U^0$. Hence we have shown that $\operatorname{span}(\varphi_3, \varphi_4, \varphi_5) \subseteq U^0$.

Example 3.123
continued

To show the inclusion in the other direction, suppose that $\varphi \in U^0$. Because the dual basis is a basis of $(\mathbb{R}^5)'$, there exist $c_1, c_2, c_3, c_4, c_5 \in \mathbb{R}$ 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 $U^0 \subseteq \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$.

Thus $U^0 = \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$.

The annihilator is a subspace Result 3.124

Suppose $U \subseteq V$. Then U^0 is a subspace of V'.

Proof. Note that $0 \in U^0$ (here 0 is the zero linear functional on V) because the zero linear functional applied to every vector in U equals $0 \in \mathbb{F}$.

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 two-dimensional subspace of $\mathbb{R}^5$, then U^0 is a three-dimensional subspace of $(\mathbb{R}^5)'$, as in Example 3.123.

The next result can be proved following the pattern of Example 3.123: 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 that $\varphi_{m+1}, \dots, \varphi_n$ is a basis of U^0, which implies the desired result. You should construct the proof just outlined, even though a slicker proof is presented here.

Dimension of the annihilator Result 3.125

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^0 = \dim V - \dim U.$$

Proof. Let $i \in \operatorname{Hom}(U, V)$ be the inclusion map defined by $i(u) = u$ for each $u \in U$. Thus i' is a linear map from V' to U'. The fundamental theorem of linear maps (3.21) applied to i' shows that

$$\dim \operatorname{im} i' + \dim \ker i' = \dim V'.$$

However, $\ker i' = U^0$ (as can be seen by thinking about the definitions) and $\dim V' = \dim V$ (by 3.111), so we can rewrite the equation above as

$$\dim \operatorname{im} i' + \dim U^0 = \dim V. \tag{3.126}$$

If $\varphi \in U'$, then φ can be extended to a linear functional ψ on V (see, for example, Exercise 13 in Section 3A). The definition of i' shows that $i'(\psi) = \varphi$. Thus $\varphi \in \operatorname{im} i'$, which implies that $\operatorname{im} i' = U'$. Hence

$$\dim \operatorname{im} i' = \dim U' = \dim U,$$

and then 3.126 becomes the equation $\dim U + \dim U^0 = \dim V$, as desired. □

The next result can be a useful tool to show that a subspace is as big as possible—see (a)—or to show that a subspace is as small as possible—see (b).

Condition for the annihilator to equal $\{0\}$ or the whole space **Result 3.127**

Suppose V is finite-dimensional and U is a subspace of V. Then

(a) $U^0 = \{0\} \iff U = V$;

(b) $U^0 = V' \iff U = \{0\}$.

Proof. To prove (a), we have

$$\begin{aligned} U^0 = \{0\} &\iff \dim U^0 = 0 \\ &\iff \dim U = \dim V \\ &\iff U = V, \end{aligned}$$

where the second equivalence follows from 3.125 and the third equivalence follows from 2.39.

Similarly, to prove (b) we have

$$\begin{aligned} U^0 = V' &\iff \dim U^0 = \dim V' \\ &\iff \dim U^0 = \dim V \\ &\iff \dim U = 0 \\ &\iff U = \{0\}, \end{aligned}$$

where one direction of the first equivalence follows from 2.39, the second equivalence follows from 3.111, and the third equivalence follows from 3.125. □

The proof of (a) in the next result does not use the hypothesis that V and W are finite-dimensional.

The null space of T' **Result 3.128**

Suppose V and W are finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Then

(a) $\ker T' = (\operatorname{im} T)^0$;

(b) $\dim \ker T' = \dim \ker T + \dim W - \dim V$.

Proof.

(a) First suppose $\varphi \in \ker T'$. Thus $0 = T'(\varphi) = \varphi \circ T$. Hence

$$0 = (\varphi \circ T)(v) = \varphi(Tv) \quad \text{for every } v \in V.$$

Thus $\varphi \in (\operatorname{im} T)^0$. This implies that $\ker T' \subseteq (\operatorname{im} T)^0$.

To prove the inclusion in the opposite direction, now suppose $\varphi \in (\operatorname{im} T)^0$. Thus $\varphi(Tv) = 0$ for every vector $v \in V$. Hence $0 = \varphi \circ T = T'(\varphi)$. In other words, $\varphi \in \ker T'$, which shows that $(\operatorname{im} T)^0 \subseteq \ker T'$, completing the proof of (a).

(b) We have

$$\begin{aligned}\dim \ker T' &= \dim(\operatorname{im} T)^0 \\ &= \dim W - \dim \operatorname{im} T \\ &= \dim W - (\dim V - \dim \ker T) \\ &= \dim \ker T + \dim W - \dim V,\end{aligned}$$

where the first equality comes from (a), the second equality comes from 3.125, and the third equality comes from the fundamental theorem of linear maps (3.21). □

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.

T surjective is equivalent to T' injective **Result 3.129**

Suppose V and W are finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Then

$$T \text{ is surjective} \iff T' \text{ is injective.}$$

Proof. We have

$$\begin{aligned}T \in \operatorname{Hom}(V, W) \text{ is surjective} &\iff \operatorname{im} T = W \\ &\iff (\operatorname{im} T)^0 = \{0\} \\ &\iff \ker T' = \{0\} \\ &\iff T' \text{ is injective,}\end{aligned}$$

where the second equivalence comes from 3.127(a) and the third equivalence comes from 3.128(a). □

The range of T' **Result 3.130**

Suppose V and W are finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Then

(a) $\dim \operatorname{im} T' = \dim \operatorname{im} T$;

(b) $\operatorname{im} T' = (\ker T)^0$.

Proof.

(a) We have

$$\begin{aligned}\dim \operatorname{im} T' &= \dim W' - \dim \ker T' \\ &= \dim W - \dim(\operatorname{im} T)^0 \\ &= \dim \operatorname{im} T,\end{aligned}$$

where the first equality comes from 3.21, the second equality comes from 3.111 and 3.128(a), and the third equality comes from 3.125.

(b) First suppose $\varphi \in \operatorname{im} T'$. Thus there exists $\psi \in W'$ such that $\varphi = T'(\psi)$. If $v \in \ker T$, then

$$\varphi(v) = \big(T'(\psi)\big)v = (\psi \circ T)(v) = \psi(Tv) = \psi(0) = 0.$$

Hence $\varphi \in (\ker T)^0$. This implies that $\operatorname{im} T' \subseteq (\ker T)^0$.

We will complete the proof by showing that $\operatorname{im} T'$ and $(\ker T)^0$ have the same dimension. To do this, note that

$$\begin{aligned}\dim \operatorname{im} T' &= \dim \operatorname{im} T \\ &= \dim V - \dim \ker T \\ &= \dim(\ker T)^0,\end{aligned}$$

where the first equality comes from (a), the second equality comes from 3.21, and the third equality comes from 3.125. □

The next result should be compared to 3.129.

T injective is equivalent to T' surjective **Result 3.131**

Suppose V and W are finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Then

$$T \text{ is injective} \iff T' \text{ is surjective.}$$

Proof. We have

$$\begin{aligned}T \text{ is injective} &\iff \ker T = \{0\} \\ &\iff (\ker T)^0 = V' \\ &\iff \operatorname{im} T' = V',\end{aligned}$$

where the second equivalence follows from 3.127(b) and the third equivalence follows from 3.130(b). □

Matrix of Dual of Linear Map

The setting for the next result is the assumption that we have a basis $v_1, \dots, v_n$ of V, along with its dual basis $\varphi_1, \dots, \varphi_n$ of V'. We also have a basis $w_1, \dots, w_m$ of W, along with its dual basis $\psi_1, \dots, \psi_m$ of W'. Thus $\operatorname{Mat}(T)$ is computed with respect to the bases just mentioned of V and W, and $\operatorname{Mat}(T')$ is computed with respect to the dual bases just mentioned of W' and V'. Using these bases gives the following pretty result.

Matrix of T' is transpose of matrix of T **Result 3.132**

Suppose V and W are finite-dimensional and $T \in \text{Hom}(V, W)$. Then

$$\text{Mat}(T') = \big(\text{Mat}(T)\big)^t.$$

Proof. Let $A = \text{Mat}(T)$ and $C = \text{Mat}(T')$. Suppose $1 \le j \le m$ and $1 \le k \le n$.

From the definition of $\text{Mat}(T')$ 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 $C = A^t$. In other words, $\text{Mat}(T') = \big(\text{Mat}(T)\big)^t$, as desired. □

Now we use duality to give an alternative proof that the column rank of a matrix equals the row rank of the matrix. This result was previously proved using different tools—see 3.57.

Column rank equals row rank **Result 3.133**

Suppose $A \in \mathbb{F}^{m,n}$. Then the column rank of A equals the row rank of A.

Proof. Define $T: \mathbb{F}^{n,1} \to \mathbb{F}^{m,1}$ by $Tx = Ax$. Thus $\text{Mat}(T) = A$, where $\text{Mat}(T)$ is computed with respect to the standard bases of $\mathbb{F}^{n,1}$ and $\mathbb{F}^{m,1}$. Now

$$\begin{aligned}\text{column rank of } A &= \text{column rank of } \text{Mat}(T)\\ &= \dim \text{im}\, T\\ &= \dim \text{im}\, T'\\ &= \text{column rank of } \text{Mat}(T')\\ &= \text{column rank of } A^t\\ &= \text{row rank of } A,\end{aligned}$$

where the second equality comes from 3.78, the third equality comes from 3.130(a), the fourth equality comes from 3.78, the fifth equality comes from 3.132, and the last equality follows from the definitions of row and column rank. □

See Exercise 8 in Section 7A for another alternative proof of the result above.

Exercises

3F.1 Explain why each linear functional is surjective or is the zero map.

3F.2 Give three distinct examples of linear functionals on $\mathbb{R}^{[0,1]}$.

3F.3 Suppose V is finite-dimensional and $v \in V$ with $v \neq 0$. Prove that there exists $\varphi \in V'$ such that $\varphi(v) = 1$.

3F.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 $\varphi \neq 0$.

3F.5 Suppose $T \in \operatorname{Hom}(V, W)$ and $w_1, \dots, w_m$ is a basis of $\operatorname{im} T$. Hence for each $v \in V$, there exist unique numbers $\varphi_1(v), \dots, \varphi_m(v)$ such that

$$Tv = \varphi_1(v)w_1 + \cdots + \varphi_m(v)w_m,$$

thus defining functions $\varphi_1, \dots, \varphi_m$ from V to $\mathbb{F}$. Show that each of the functions $\varphi_1, \dots, \varphi_m$ is a linear functional on V.

3F.6 Suppose $\varphi, \beta \in V'$. Prove that $\ker \varphi \subseteq \ker \beta$ if and only if there exists $c \in \mathbb{F}$ such that $\beta = c\varphi$.

3F.7 Suppose that $V_1, \dots, V_m$ are vector spaces. Prove that $(V_1 \times \cdots \times V_m)'$ and ${V_1}' \times \cdots \times {V_m}'$ are isomorphic vector spaces.

3F.8 Suppose $v_1, \dots, v_n$ is a basis of V and $\varphi_1, \dots, \varphi_n$ is the dual basis of V'. Define $\Gamma \colon V \to \mathbb{F}^n$ and $\Lambda \colon \mathbb{F}^n \to V$ by

$$\Gamma(v) = \big(\varphi_1(v), \dots, \varphi_n(v)\big) \quad \text{and} \quad \Lambda(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

Explain why Γ and Λ are inverses of each other.

3F.9 Suppose m is a positive integer. Show that the dual basis of the basis $1, x, \dots, x^m$ of $\mathbb{R}[X]_{\leq m}$ is $\varphi_0, \varphi_1, \dots, \varphi_m$, where

$$\varphi_k(p) = \frac{p^{(k)}(0)}{k!}.$$

Here $p^{(k)}$ denotes the kth derivative of p, with the understanding that the 0th derivative of p is p.

3F.10 Suppose m is a positive integer.
(a) Show that $1, x-5, \ldots, (x-5)^m$ is a basis of $\mathbb{R}[X]_{\le m}$.

(b) What is the dual basis of the basis in (a)?

3F.11 Suppose $v_1, \ldots, v_n$ is a basis of V and $\varphi_1, \ldots, \varphi_n$ is the corresponding dual basis of V'. Suppose $\psi \in V'$. Prove that

$$\psi = \psi(v_1)\varphi_1 + \cdots + \psi(v_n)\varphi_n.$$

3F.12 Suppose $S, T \in \text{Hom}(V, W)$.
(a) Prove that $(S+T)' = S' + T'$.

(b) Prove that $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbb{F}$.

This exercise asks you to verify (a) *and* (b) *in 3.120.*

3F.13 Show that the dual map of the identity operator on V is the identity operator on V'.

3F.14 Define $T: \mathbb{R}^3 \to \mathbb{R}^2$ 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 $\mathbb{R}^2$ and ψ_1, ψ_2, ψ_3 denotes the dual basis of the standard basis of $\mathbb{R}^3$.
(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.

3F.15 Define $T: \mathbb{R}[X] \to \mathbb{R}[X]$ by

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

for each $x \in \mathbb{R}$.
(a) Suppose $\varphi \in \mathbb{R}[X]'$ is defined by $\varphi(p) = p'(4)$. Describe the linear functional $T'(\varphi)$ on $\mathbb{R}[X]$.

(b) Suppose $\varphi \in \mathbb{R}[X]'$ is defined by $\varphi(p) = \int_0^1 p$. Evaluate $\big(T'(\varphi)\big)(x^3)$.

3F.16 Suppose W is finite-dimensional and $T \in \text{Hom}(V, W)$. Prove that

$$T' = 0 \iff T = 0.$$

3F.17 Suppose V and W are finite-dimensional and $T \in \text{Hom}(V, W)$. Prove that T is invertible if and only if $T' \in \text{Hom}(W', V')$ is invertible.

3F.18 Suppose V and W are finite-dimensional. Prove that the map that takes $T \in \text{Hom}(V, W)$ to $T' \in \text{Hom}(W', V')$ is an isomorphism of $\text{Hom}(V, W)$ onto $\text{Hom}(W', V')$.

3F.19 Suppose $U \subseteq V$. Explain why

$$U^0 = \{\varphi \in V' : U \subseteq \ker \varphi\}.$$

3F.20 Suppose V is finite-dimensional and U is a subspace of V. Show that

$$U = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in U^0\}.$$

3F.21 Suppose V is finite-dimensional and U and W are subspaces of V.

(a) Prove that $W^0 \subseteq U^0$ if and only if $U \subseteq W$.

(b) Prove that $W^0 = U^0$ if and only if $U = W$.

3F.22 Suppose V is finite-dimensional and U and W are subspaces of V.

(a) Show that $(U + W)^0 = U^0 \cap W^0$.

(b) Show that $(U \cap W)^0 = U^0 + W^0$.

3F.23 Suppose V is finite-dimensional and $\varphi_1, \ldots, \varphi_m \in V'$. Prove that the following three sets are equal to each other.

(a) $\text{span}(\varphi_1, \ldots, \varphi_m)$

(b) $\big((\ker \varphi_1) \cap \cdots \cap (\ker \varphi_m)\big)^0$

(c) $\{\varphi \in V' : (\ker \varphi_1) \cap \cdots \cap (\ker \varphi_m) \subseteq \ker \varphi\}$

3F.24 Suppose V is finite-dimensional and $v_1, \ldots, v_m \in V$. Define a linear map $\Gamma \colon V' \to \mathbb{F}^m$ by $\Gamma(\varphi) = \big(\varphi(v_1), \ldots, \varphi(v_m)\big)$.

(a) Prove that $v_1, \ldots, v_m$ spans V if and only if Γ is injective.

(b) Prove that $v_1, \ldots, v_m$ is linearly independent if and only if Γ is surjective.

3F.25 Suppose V is finite-dimensional and $\varphi_1, \ldots, \varphi_m \in V'$. Define a linear map $\Gamma \colon V \to \mathbb{F}^m$ by $\Gamma(v) = \big(\varphi_1(v), \ldots, \varphi_m(v)\big)$.

(a) Prove that $\varphi_1, \ldots, \varphi_m$ spans V' if and only if Γ is injective.

(b) Prove that $\varphi_1, \ldots, \varphi_m$ is linearly independent if and only if Γ is surjective.

3F.26 Suppose V is finite-dimensional and Ω is a subspace of V'. Prove that

$$\Omega = \{v \in V : \varphi(v) = 0 \text{ for every } \varphi \in \Omega\}^0.$$

3F.27 Suppose $T \in \text{End}\big(\mathbb{R}[X]_{\le 5}\big)$ and $\ker T' = \text{span}(\varphi)$, where φ is the linear functional on $\mathbb{R}[X]_{\le 5}$ defined by $\varphi(p) = p(8)$. Prove that

$$\text{im}\, T = \{p \in \mathbb{R}[X]_{\le 5} : p(8) = 0\}.$$

3F.28 Suppose V is finite-dimensional and $\varphi_1, \dots, \varphi_m$ is a linearly independent list in V'. Prove that

$$\dim\big((\ker \varphi_1) \cap \cdots \cap (\ker \varphi_m)\big) = (\dim V) - m.$$

3F.29 Suppose V and W are finite-dimensional and $T \in \text{Hom}(V, W)$.

(a) Prove that if $\varphi \in W'$ and $\ker T' = \text{span}(\varphi)$, then $\text{im}\, T = \ker \varphi$.

(b) Prove that if $\psi \in V'$ and $\text{im}\, T' = \text{span}(\psi)$, then $\ker T = \ker \psi$.

3F.30 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 $\varphi_1, \dots, \varphi_n$.

3F.31 Suppose U is a subspace of V. Let $i: U \to V$ be the inclusion map defined by $i(u) = u$. Thus $i' \in \text{Hom}(V', U')$.

(a) Show that $\ker i' = U^0$.

(b) Prove that if V is finite-dimensional, then $\text{im}\, i' = U'$.

(c) Prove that if V is finite-dimensional, then $\tilde{i}'$ is an isomorphism from V'/U^0 onto U'.

The isomorphism in (c) *is natural in that it does not depend on a choice of basis in either vector space.*

3F.32 The *double dual space* of V, denoted by V'', is defined to be the dual space of V'. In other words, $V'' = (V')'$. Define $\Lambda: V \to V''$ by

$$(\Lambda v)(\varphi) = \varphi(v)$$

for each $v \in V$ and each $\varphi \in V'$.

(a) Show that Λ is a linear map from V to V''.

(b) Show that if $T \in \text{End}(V)$, 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.

3F.33 Suppose U is a subspace of V. Let $\pi: V \to V/U$ be the usual quotient map. Thus $\pi' \in \text{Hom}\big((V/U)', V'\big)$.

(a) Show that π' is injective.

(b) Show that $\text{im}\, \pi' = U^0$.

(c) Conclude that π' is an isomorphism from $(V/U)'$ onto U^0.

The isomorphism in (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.*

Part 2



Eigenvalues

CHAPTER 4	Polynomials	121
CHAPTER 5	Eigenvalues and Eigenvectors	133

4 Polynomials

This chapter contains material on polynomials that we will use to investigate linear maps from a vector space to itself. Many 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 results in this chapter—they will be used in later chapters.

This chapter begins with a brief discussion of algebraic properties of the complex numbers. Then we prove that a nonconstant polynomial cannot have more zeros than its degree. We also give a linear-algebra-based proof of the division algorithm for polynomials, which is worth reading even if you are already familiar with a proof that does not use linear algebra.

As we will see, the fundamental theorem of algebra leads to a factorization of every polynomial into degree-one factors if the scalar field is $\mathbb{C}$ or to factors of degree at most two if the scalar field is $\mathbb{R}$.

Standing assumption for this chapter — Definition

- $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$.

Before discussing polynomials with complex or real coefficients, we need to learn a bit more about the complex numbers.

Real part, $\operatorname{Re} z$, imaginary part, $\operatorname{Im} z$ — Definition 4.1

Suppose $z = a + bi$, where a and b are real numbers.

- The *real part* of z, denoted by $\operatorname{Re} z$, is defined by $\operatorname{Re} z = a$.
- The *imaginary part* of z, denoted by $\operatorname{Im} z$, is defined by $\operatorname{Im} z = b$.

Thus for every complex number z, we have

$$z = \operatorname{Re} z + (\operatorname{Im} z)i.$$

Definition 4.2 **Complex conjugate, $\overline{z}$, absolute value, $|z|$**

Suppose $z \in \mathbb{C}$.

- The *complex conjugate* of $z \in \mathbb{C}$, denoted by $\overline{z}$, is defined by

$$\overline{z} = \operatorname{Re} z - (\operatorname{Im} z)i.$$

- The *absolute value* of a complex number z, denoted by $|z|$, is defined by

$$|z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

Example 4.3 **Real and imaginary part, complex conjugate, absolute value**

Suppose $z = 3 + 2i$. Then

- $\operatorname{Re} z = 3$ and $\operatorname{Im} z = 2$;
- $\overline{z} = 3 - 2i$;
- $|z| = \sqrt{3^2 + 2^2} = \sqrt{13}$.

Identifying a complex number $z \in \mathbb{C}$ with the ordered pair $(\operatorname{Re} z, \operatorname{Im} z) \in \mathbb{R}^2$ identifies $\mathbb{C}$ with $\mathbb{R}^2$. Note that $\mathbb{C}$ is a one-dimensional complex vector space, but we can also think of $\mathbb{C}$ (identified with $\mathbb{R}^2$) as a two-dimensional real vector space.

The absolute value of each complex number is a nonnegative number. Specifically, if $z \in \mathbb{C}$, then $|z|$ equals the distance from the origin in $\mathbb{R}^2$ to the point $(\operatorname{Re} z, \operatorname{Im} z) \in \mathbb{R}^2$.

You should verify that $z = \overline{z}$ if and only if z is a real number.

The real and imaginary parts, complex conjugate, and absolute value have the properties listed in the following multipart result.

Result 4.4 **Properties of complex numbers**

Suppose $w, z \in \mathbb{C}$. Then the following equalities and inequalities hold.

sum of z and $\overline{z}$
$z + \overline{z} = 2 \operatorname{Re} z$.

difference of z and $\overline{z}$
$z - \overline{z} = 2(\operatorname{Im} z)i$.

product of z and $\overline{z}$
$z\overline{z} = |z|^2$.

additivity and multiplicativity of complex conjugate
$\overline{w + z} = \overline{w} + \overline{z}$ and $\overline{wz} = \overline{w}\,\overline{z}$.

double complex conjugate
$\overline{\overline{z}} = z$.

real and imaginary parts are bounded by $|z|$
$|\operatorname{Re} z| \le |z|$ and $|\operatorname{Im} z| \le |z|$.

Result 4.4
continued

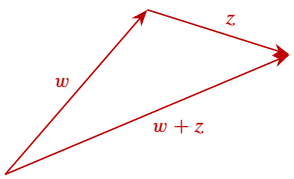
absolute value of the complex conjugate
$|\overline{z}| = |z|$.
multiplicativity of absolute value
$|wz| = |w|\,|z|$.
triangle inequality
$|w + z| \le |w| + |z|$.

Proof. Except for the last item above, the routine verifications of the assertions above are left to the reader. To verify the triangle inequality, we have

See Exercise 2 for the reverse triangle inequality.

$$\begin{aligned}
|w+z|^2 &= (w+z)(\overline{w}+\overline{z}) \\
&= w\overline{w} + z\overline{z} + w\overline{z} + z\overline{w} \\
&= |w|^2 + |z|^2 + w\overline{z} + \overline{w\overline{z}} \\
&= |w|^2 + |z|^2 + 2\,\mathrm{Re}(w\overline{z}) \\
&\le |w|^2 + |z|^2 + 2|w\overline{z}| \\
&= |w|^2 + |z|^2 + 2|w|\,|z| \\
&= (|w| + |z|)^2.
\end{aligned}$$

Taking square roots now gives the desired inequality $|w + z| \le |w| + |z|$. □



Geometric interpretation of triangle inequality: The length of each side of a triangle is less than or equal to the sum of the lengths of the two other sides.

Zeros of Polynomials

Recall that a function $p: \mathbb{F} \to \mathbb{F}$ is called a polynomial of degree m if there exist $a_0, \ldots, a_m \in \mathbb{F}$ with $a_m \neq 0$ such that

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

for all $z \in \mathbb{F}$. A polynomial could have more than one degree if the representation of p in the form above were not unique. Our first task is to show that this cannot happen.

The solutions to the equation $p(z) = 0$ play a crucial role in the study of a polynomial $p \in \mathbb{F}[X]$. Thus these solutions have a special name.

Zero of a polynomial Definition 4.5
A number $\lambda \in \mathbb{F}$ is called a *zero* (or *root*) of a polynomial $p \in \mathbb{F}[X]$ if

$$p(\lambda) = 0.$$

The next result is the key tool that we will use to show that the degree of a polynomial is unique.

Each zero of a polynomial corresponds to a degree-one factor Result 4.6

Suppose m is a positive integer and $p \in \mathbb{F}[X]$ is a polynomial of degree m. Suppose $\lambda \in \mathbb{F}$. Then $p(\lambda) = 0$ if and only if there exists a polynomial $q \in \mathbb{F}[X]$ of degree $m - 1$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbb{F}$.

Proof. First suppose $p(\lambda) = 0$. Let $a_0, a_1, \ldots, a_m \in \mathbb{F}$ be such that

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

for all $z \in \mathbb{F}$. Then

$$p(z) = p(z) - p(\lambda) = a_1(z - \lambda) + \cdots + a_m(z^m - \lambda^m) \tag{4.7}$$

for all $z \in \mathbb{F}$. For each $k \in \{1, \ldots, m\}$, the equation

$$z^k - \lambda^k = (z - \lambda) \sum_{j=1}^{k} \lambda^{j-1} z^{k-j}$$

shows that $z^k - \lambda^k$ equals $z - \lambda$ times some polynomial of degree $k - 1$. Thus 4.7 shows that p equals $z - \lambda$ times some polynomial of degree $m - 1$, as desired.

To prove the implication in the other direction, now suppose that there is a polynomial $q \in \mathbb{F}[X]$ such that $p(z) = (z - \lambda)q(z)$ for every $z \in \mathbb{F}$. Then $p(\lambda) = (\lambda - \lambda)q(\lambda) = 0$, as desired. □

Now we can prove that polynomials do not have too many zeros.

Degree m implies at most m zeros Result 4.8

Suppose m is a positive integer and $p \in \mathbb{F}[X]$ is a polynomial of degree m. Then p has at most m zeros in $\mathbb{F}$.

Proof. We will use induction on m. The desired result holds if $m = 1$ because if $a_1 \neq 0$ then the polynomial $a_0 + a_1 z$ has only one zero (which equals $-a_0/a_1$). Thus assume that $m > 1$ and the desired result holds for $m - 1$.

If p has no zeros in $\mathbb{F}$, then the desired result holds and we are done. Thus suppose p has a zero $\lambda \in \mathbb{F}$. By 4.6, there is polynomial $q \in \mathbb{F}[X]$ of degree $m - 1$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbb{F}$. Our induction hypothesis implies that q has at most $m - 1$ zeros in $\mathbb{F}$. The equation above shows that the zeros of p in $\mathbb{F}$ are exactly the zeros of q in $\mathbb{F}$ along with λ. Thus p has at most m zeros in $\mathbb{F}$. □

The 0 polynomial is declared to have degree $-\infty$ so that exceptions are not needed for various reasonable results such as $\deg(pq) = \deg p + \deg q$.

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 polynomial with some nonzero

coefficients but infinitely many zeros). In particular, the degree of a polynomial is uniquely defined.

Recall that the degree of the 0 polynomial is defined to be $-\infty$. When necessary, use the expected arithmetic with $-\infty$. For example, $-\infty < m$ and $-\infty + m = -\infty$ for every integer m.

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 result gives an analogous result for polynomials. Thus the next result 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 a remainder polynomial r when the polynomial p is divided by the polynomial s.

The division algorithm for polynomials could be proved without using any linear algebra. However, as is appropriate for a linear algebra textbook, the proof given here uses linear algebra techniques and makes nice use of a basis of $\mathbb{F}[X]_{\leq n}$, which is the $(n+1)$-dimensional vector space of polynomials with coefficients in $\mathbb{F}$ and of degree at most n.

Division algorithm for polynomials **Result 4.9**

Suppose that $p, s \in \mathbb{F}[X]$, with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathbb{F}[X]$ such that

$$p = sq + r$$

and $\deg r < \deg s$.

Proof. Let $n = \deg p$ and let $m = \deg s$. If $n < m$, then take $q = 0$ and $r = p$ to get the desired equation $p = sq + r$ with $\deg r < \deg s$. Thus we now assume that $n \geq m$.

The list

$$1, z, \ldots, z^{m-1}, s, zs, \ldots, z^{n-m}s \tag{4.10}$$

is linearly independent in $\mathbb{F}[X]_{\leq n}$ because each polynomial in this list has a different degree. Also, the list 4.10 has length $n+1$, which equals $\dim \mathbb{F}[X]_{\leq n}$. Hence 4.10 is a basis of $\mathbb{F}[X]_{\leq n}$ [by 2.38].

Because $p \in \mathbb{F}[X]_{\leq n}$ and 4.10 is a basis of $\mathbb{F}[X]_{\leq n}$, there exist unique constants $a_0, a_1, \ldots, a_{m-1} \in \mathbb{F}$ and $b_0, b_1, \ldots, b_{n-m} \in \mathbb{F}$ such that

$$\begin{aligned} p &= a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + b_0 s + b_1 z s + \cdots + b_{n-m} z^{n-m} s \\ &= \underbrace{a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}}_{r} + s(\underbrace{b_0 + b_1 z + \cdots + b_{n-m} z^{n-m}}_{q}). \end{aligned} \tag{4.11}$$

With r and q as defined above, we see that p can be written as $p = sq + r$ with $\deg r < \deg s$, as desired.

The uniqueness of $q, r \in \mathbb{F}[X]$ satisfying these conditions follows from the uniqueness of the constants $a_0, a_1, \ldots, a_{m-1} \in \mathbb{F}$ and $b_0, b_1, \ldots, b_{n-m} \in \mathbb{F}$ satisfying 4.11. □

Factorization of Polynomials over $\mathbb{C}$

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.

We have been handling polynomials with complex coefficients and polynomials with real coefficients simultaneously, letting $\mathbb{F}$ denote $\mathbb{R}$ or $\mathbb{C}$. Now we will see differences between these two cases. First we treat polynomials with complex coefficients. Then we will use those results to prove corresponding results for polynomials with real coefficients.

Our proof of the fundamental theorem of algebra implicitly uses the result that a continuous real-valued function on a closed disk in $\mathbb{R}^2$ attains a minimum value. A web search can lead you to several other proofs of the fundamental theorem of algebra. The proof using Liouville's theorem is particularly nice if you are comfortable with analytic functions. All proofs of the fundamental theorem of algebra need to use some analysis, because the result is not true if $\mathbb{C}$ is replaced, for example, with the set of numbers of the form $c + di$ where c, d are rational numbers.

Result 4.12

Fundamental theorem of algebra, first version
Every nonconstant polynomial with complex coefficients has a zero in $\mathbb{C}$.

Proof. De Moivre's theorem, which you can prove using induction on k and the addition formulas for cosine and sine, states that if k is a positive integer and $\theta \in \mathbb{R}$, then

$$(\cos\theta + i\sin\theta)^k = \cos k\theta + i\sin k\theta.$$

Suppose $w \in \mathbb{C}$ and k is a positive integer. Using polar coordinates, we know that there exist $r \geq 0$ and $\theta \in \mathbb{R}$ such that

$$r(\cos\theta + i\sin\theta) = w.$$

De Moivre's theorem implies that

$$\left(r^{1/k}(\cos\tfrac{\theta}{k} + i\sin\tfrac{\theta}{k})\right)^k = w.$$

Thus every complex number has a kth root, a fact that we will soon use.

Suppose p is a nonconstant polynomial with complex coefficients and highest-order nonzero term $c_m z^m$. Then $|p(z)| \to \infty$ as $|z| \to \infty$ (because $|p(z)|/|z^m| \to |c_m|$ as $|z| \to \infty$). Thus the continuous function $z \mapsto |p(z)|$ has a global minimum at some point $\zeta \in \mathbb{C}$. To show that $p(\zeta) = 0$, suppose that $p(\zeta) \neq 0$.

Define a new polynomial q by

$$q(z) = \frac{p(z+\zeta)}{p(\zeta)}.$$

The function $z \mapsto |q(z)|$ has a global minimum value of 1 at $z = 0$. Write

$$q(z) = 1 + a_k z^k + \cdots + a_m z^m,$$

where k is the smallest positive integer such that the coefficient of z^k is nonzero; in other words, $a_k \neq 0$.

Let $\beta \in \mathbb{C}$ be such that $\beta^k = -1/a_k$. There is a constant $c > 1$ such that if $t \in (0, 1)$, then

$$\begin{aligned} |q(t\beta)| &\leq |1 + a_k t^k \beta^k| + t^{k+1} c \\ &= 1 - t^k (1 - tc). \end{aligned}$$

Thus taking t to be $1/(2c)$ in the inequality above, we have $|q(t\beta)| < 1$, which contradicts the assumption that the global minimum of $z \mapsto |q(z)|$ is 1. This contradiction implies that $p(\zeta) = 0$, showing that p has a zero, as desired. □

Computers 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 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 can find that the zeros of p are approximately the five numbers -1.87, -0.74, 0.62, 1.47, 5.51.

The first version of the fundamental theorem of algebra leads to the following factorization result for polynomials with complex coefficients. Note that in this factorization, the zeros of p are the numbers $\lambda_1, \ldots, \lambda_m$, which are the only values of z for which the right side of the equation in the next result equals 0.

Fundamental theorem of algebra, second version **Result 4.13**

If $p \in \mathbb{C}[X]$ 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 $c, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$.

Proof. Let $p \in \mathbb{C}[X]$ and let $m = \deg p$. We will use induction on m. If $m = 1$, then 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 first version of the fundamental theorem of algebra (4.12), p has a zero $\lambda \in \mathbb{C}$. By 4.6, there is a polynomial q of degree $m - 1$ such that

$$p(z) = (z - \lambda) q(z)$$

for all $z \in \mathbb{C}$. 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. The number c is uniquely determined as the coefficient of z^m in p. So we only need to 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 $z \in \mathbb{C}$, 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 if $z \neq \lambda_1$, 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 $z \in \mathbb{C}$ except possibly $z = \lambda_1$. Actually the equation above holds for all $z \in \mathbb{C}$, 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. □

The failure of the fundamental theorem of algebra for $\mathbb{R}$ accounts for the differences between linear algebra on real and complex vector spaces, as we will see in later chapters.

Factorization of Polynomials over $\mathbb{R}$

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 $\mathbb{R}$, we will use our factorization theorem over $\mathbb{C}$. We begin with the next result.

Polynomials with real coefficients have nonreal zeros in pairs **Result 4.14**

Suppose $p \in \mathbb{C}[X]$ is a polynomial with real coefficients. If $\lambda \in \mathbb{C}$ is a zero of p, then so is $\overline{\lambda}$.

Proof. Let

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

where $a_0, \dots, a_m$ are real numbers. Suppose $\lambda \in \mathbb{C}$ 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

$$a_0 + a_1 \overline{\lambda} + \cdots + a_m \overline{\lambda}^m = 0,$$

where we have used basic properties of the complex conjugate (see 4.4). The equation above shows that $\overline{\lambda}$ is a zero of p. □

We want a factorization theorem for polynomials with real coefficients. We begin with the following result.

Think about the quadratic formula in connection with the result below.

Factorization of a quadratic polynomial **Result 4.15**

Suppose $b, c \in \mathbb{R}$. Then there is a polynomial factorization of the form

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

with $\lambda_1, \lambda_2 \in \mathbb{R}$ if and only if $b^2 \geq 4c$.

Proof. Notice that

$$x^2 + bx + c = \Big(x + \frac{b}{2}\Big)^2 + \Big(c - \frac{b^2}{4}\Big).$$

The equation above is the basis of the technique called completing the square.

First suppose $b^2 < 4c$. Then the right side of the equation above is positive for every $x \in \mathbb{R}$. 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 $\lambda_1, \lambda_2 \in \mathbb{R}$.

Conversely, now suppose $b^2 \geq 4c$. Then there is a real number d such that $d^2 = b^2/4 - c$. From the displayed equation above, we have

$$\begin{aligned} x^2 + bx + c &= \Big(x + \frac{b}{2}\Big)^2 - d^2 \\ &= \Big(x + \frac{b}{2} + d\Big)\Big(x + \frac{b}{2} - d\Big), \end{aligned}$$

which gives the desired factorization. □

The next result gives a factorization of a polynomial over $\mathbb{R}$. The idea of the proof is to use the second version of the fundamental theorem of algebra (4.13), which gives a factorization of p as a polynomial with complex coefficients. Complex but nonreal zeros of p come in pairs; see 4.14. Thus if the factorization of p as an element of $\mathbb{C}[X]$ includes terms of the form $(x - \lambda)$ with λ a nonreal complex number, then $(x - \overline{\lambda})$ is also a term in the factorization. Multiplying together these two terms, we get

$$x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2,$$

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 $\mathbb{C}[X]$. We are guaranteed by 4.14 that $(x - \overline{\lambda})$ also appears as a term in the factorization, but 4.14 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.

Factorization of a polynomial over $\mathbb{R}$ **Result 4.16**

Suppose $p \in \mathbb{R}[X]$ is a nonconstant polynomial. Then p has a unique factoriza-

Result 4.16 *continued*

tion (except for the order of the factors) of the form

$$p(x) = c(x-\lambda_1)\cdots(x-\lambda_m)(x^2+b_1x+c_1)\cdots(x^2+b_Mx+c_M),$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbb{R}$, with $b_k^2 < 4c_k$ for each k.

Proof. First we will prove that the desired factorization exists, and after that we will prove the uniqueness.

Think of p as an element of $\mathbb{C}[X]$. If all (complex) zeros of p are real, then we have the desired factorization by 4.13. Thus suppose p has a zero $\lambda \in \mathbb{C}$ with $\lambda \notin \mathbb{R}$. By 4.14, $\overline{\lambda}$ is a zero of p. Thus we can write

$$\begin{aligned} p(x) &= (x-\lambda)(x-\overline{\lambda})q(x) \\ &= \big(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\big)q(x) \end{aligned}$$

for some polynomial $q \in \mathbb{C}[X]$ of degree two less than the degree of p. If we can prove that q has real coefficients, then using induction on the degree of p completes the proof of the existence part of this result.

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 $x \in \mathbb{R}$. The equation above implies that $q(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$. Writing

$$q(x) = a_0 + a_1x + \cdots + a_{n-2}x^{n-2},$$

where $n = \deg p$ and $a_0, \ldots, a_{n-2} \in \mathbb{C}$, 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 $x \in \mathbb{R}$. This implies that $\operatorname{Im} a_0, \ldots, \operatorname{Im} a_{n-2}$ all equal 0 (by 4.8). Thus all 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 $x^2 + b_kx + c_k$ with $b_k^2 < 4c_k$ can be uniquely written as $(x-\lambda_k)(x-\overline{\lambda_k})$ with $\lambda_k \in \mathbb{C}$. A moment's thought shows that two different factorizations of p as an element of $\mathbb{R}[X]$ would lead to two different factorizations of p as an element of $\mathbb{C}[X]$, contradicting 4.13. □

Exercises

4.1 Suppose $w, z \in \mathbb{C}$. Verify the following equalities and inequalities.

(a) $z + \overline{z} = 2\operatorname{Re} z$

(b) $z - \overline{z} = 2(\operatorname{Im} z)i$

(c) $z\overline{z} = |z|^2$

(d) $\overline{w+z} = \overline{w} + \overline{z}$ and $\overline{wz} = \overline{w}\,\overline{z}$

(e) $\overline{\overline{z}} = z$

(f) $|\operatorname{Re} z| \le |z|$ and $|\operatorname{Im} z| \le |z|$

(g) $|\overline{z}| = |z|$

(h) $|wz| = |w|\,|z|$

The results above are the parts of 4.4 that were left to the reader.

4.2 Prove that if $w, z \in \mathbb{C}$, then $\big||w| - |z|\big| \le |w - z|$.

The inequality above is called the ***reverse triangle inequality****.*

4.3 Suppose V is a complex vector space and $\varphi \in V'$. Define $\sigma\colon V \to \mathbb{R}$ by $\sigma(v) = \operatorname{Re}\varphi(v)$ for each $v \in V$. Show that

$$\varphi(v) = \sigma(v) - i\sigma(iv)$$

for all $v \in V$.

4.4 Suppose m is a positive integer. Is the set

$$\{0\} \cup \{p \in \mathbb{F}[X] : \deg p = m\}$$

a subspace of $\mathbb{F}[X]$?

4.5 Is the set

$$\{0\} \cup \{p \in \mathbb{F}[X] : \deg p \text{ is even}\}$$

a subspace of $\mathbb{F}[X]$?

4.6 Suppose that m and n are positive integers with $m \le n$, and suppose $\lambda_1, \dots, \lambda_m \in \mathbb{F}$. Prove that there exists a polynomial $p \in \mathbb{F}[X]$ with $\deg p = n$ such that $0 = p(\lambda_1) = \cdots = p(\lambda_m)$ and such that p has no other zeros.

4.7 Suppose that m is a nonnegative integer, $z_1, \dots, z_{m+1}$ are distinct elements of $\mathbb{F}$, and $w_1, \dots, w_{m+1} \in \mathbb{F}$. Prove that there exists a unique polynomial $p \in \mathbb{F}[X]_{\le m}$ such that

$$p(z_k) = w_k$$

for each $k = 1, \dots, m + 1$.

This result can be proved without using linear algebra. However, try to find the clearer, shorter proof that uses some linear algebra.

4.8 Suppose $p \in \mathbb{C}[X]$ has degree m. Prove that p has m distinct zeros if and only if p and its derivative p' have no zeros in common.

4.9 Prove that every polynomial of odd degree with real coefficients has a real zero.

4.10 For $p \in \mathbb{R}[X]$, define $Tp: \mathbb{R} \to \mathbb{R}$ by

$$(Tp)(x) = \begin{cases} \dfrac{p(x) - p(3)}{x - 3} & \text{if } x \neq 3, \\ p'(3) & \text{if } x = 3 \end{cases}$$

for each $x \in \mathbb{R}$. Show that $Tp \in \mathbb{R}[X]$ for every polynomial $p \in \mathbb{R}[X]$ and also show that $T: \mathbb{R}[X] \to \mathbb{R}[X]$ is a linear map.

4.11 Suppose $p \in \mathbb{C}[X]$. Define $q: \mathbb{C} \to \mathbb{C}$ by

$$q(z) = p(z)\,\overline{p(\overline{z})}.$$

Prove that q is a polynomial with real coefficients.

4.12 Suppose m is a nonnegative integer and $p \in \mathbb{C}[X]_{\leq m}$ is such that there are distinct real numbers $x_0, x_1, \ldots, x_m$ with $p(x_k) \in \mathbb{R}$ for each $k = 0, 1, \ldots, m$. Prove that all coefficients of p are real.

4.13 Suppose $p \in \mathbb{F}[X]$ with $p \neq 0$. Let $U = \{pq : q \in \mathbb{F}[X]\}$.

(a) Show that $\dim \mathbb{F}[X]/U = \deg p$.

(b) Find a basis of $\mathbb{F}[X]/U$.

4.14 Suppose $p, q \in \mathbb{C}[X]$ are nonconstant polynomials with no zeros in common. Let $m = \deg p$ and $n = \deg q$. Use linear algebra as outlined below in (a)–(c) to prove that there exist $r \in \mathbb{C}[X]_{\leq n-1}$ and $s \in \mathbb{C}[X]_{\leq m-1}$ such that

$$rp + sq = 1.$$

(a) Define $T: \mathbb{C}[X]_{\leq n-1} \times \mathbb{C}[X]_{\leq m-1} \to \mathbb{C}[X]_{\leq m+n-1}$ by

$$T(r, s) = rp + sq.$$

Show that the linear map T is injective.

(b) Show that the linear map T in (a) is surjective.

(c) Use (b) to conclude that there exist $r \in \mathbb{C}[X]_{\leq n-1}$ and $s \in \mathbb{C}[X]_{\leq m-1}$ such that $rp + sq = 1$.

5 Eigenvalues and Eigenvectors

Linear maps from one vector space to another vector space were the objects of study in Chapter 3. Now we begin our investigation of operators, which are linear maps from a vector space to itself. Their study constitutes the most important part of linear algebra.

To learn about an operator, we might try restricting it to a smaller subspace. Asking for that restriction to be an operator will lead us to the notion of invariant subspaces. Each one-dimensional invariant subspace arises from a vector that the operator maps into a scalar multiple of the vector. This path will lead us to eigenvectors and eigenvalues.

We will then prove one of the most important results in linear algebra: every operator on a finite-dimensional nonzero complex vector space has an eigenvalue. This result will allow us to show 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 at least almost half its entries equal to 0.

Standing assumptions for this chapter — **Definition**

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

5A Invariant Subspaces

Eigenvalues

Operator — **Definition 5.1**

A linear map from a vector space to itself is called an *operator*.

Suppose $T \in \text{End}(V)$. If $m \geq 2$ and

$$V = V_1 \oplus \cdots \oplus V_m,$$

Recall that we defined the notation $\text{End}(V)$ to mean $\text{Hom}(V, V)$.

where each V_k is a nonzero subspace of V, then to understand the behavior of T we only need to understand the behavior of each $T|_{V_k}$; here $T|_{V_k}$ denotes the restriction of T to the smaller domain V_k. Dealing with $T|_{V_k}$ should be easier than dealing with T because V_k 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: $T|_{V_k}$ may not map V_k into itself; in other words, $T|_{V_k}$ may not be an operator on V_k. Thus we are led to consider only decompositions of V of the form above in which T maps each V_k into itself. Hence we now give a name to subspaces of V that get mapped into themselves by T.

Invariant subspace **Definition 5.2**

Suppose $T \in \text{End}(V)$. A subspace U of V is called *invariant* under T if $Tu \in U$ for every $u \in U$.

Thus U is invariant under T if $T|_U$ is an operator on U.

Example 5.3 **Subspace invariant under differentiation operator**

Suppose that $T \in \text{End}(\mathbb{R}[X])$ is defined by $Tp = p'$. Then $\mathbb{R}[X]_{\leq 4}$, which is a subspace of $\mathbb{R}[X]$, is invariant under T because if $p \in \mathbb{R}[X]$ has degree at most 4, then p' also has degree at most 4.

Example 5.4 **Four invariant subspaces, not necessarily all different**

If $T \in \text{End}(V)$, then the following subspaces of V are all invariant under T.

- $\{0\}$ The subspace $\{0\}$ is invariant under T because if $u \in \{0\}$, then $u = 0$ and hence $Tu = 0 \in \{0\}$.
- V The subspace V is invariant under T because if $u \in V$, then $Tu \in V$.
- $\ker T$ The subspace $\ker T$ is invariant under T because if $u \in \ker T$, then $Tu = 0$, and hence $Tu \in \ker T$.
- $\text{im}\, T$ The subspace $\text{im}\, T$ is invariant under T because if $u \in \text{im}\, T$, then $Tu \in \text{im}\, T$.

Must an operator $T \in \text{End}(V)$ 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 $\mathbb{F} = \mathbb{C}$) or $\dim V > 2$ (for $\mathbb{F} = \mathbb{R}$); see 5.19 and Exercise 29 in Section 5B.

The previous example noted that $\ker T$ and $\text{im}\, T$ are invariant under T. However, these subspaces do not necessarily provide easy answers to the question above about the existence of invariant subspaces other than $\{0\}$ and V, because $\ker T$ may equal $\{0\}$ and $\text{im}\, T$ may equal V (this happens when T is invertible).

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 of dimension one.

Take any $v \in V$ with $v \neq 0$ and let U equal the set of all scalar multiples of v:

$$U = \{\lambda v : \lambda \in \mathbb{F}\} = \operatorname{span}(v).$$

Then U is a one-dimensional subspace of V (and every one-dimensional subspace of V is of this form for an appropriate choice of v). If U is invariant under an operator $T \in \operatorname{End}(V)$, then $Tv \in U$, and hence there is a scalar $\lambda \in \mathbb{F}$ such that

$$Tv = \lambda v.$$

Conversely, if $Tv = \lambda v$ for some $\lambda \in \mathbb{F}$, then $\operatorname{span}(v)$ is a one-dimensional subspace of V invariant under T.

The equation $Tv = \lambda v$, which we have just seen is intimately connected with one-dimensional invariant subspaces, is important enough that the scalars λ and vectors v satisfying it are given special names.

Eigenvalue — Definition 5.5

Suppose $T \in \operatorname{End}(V)$. A number $\lambda \in \mathbb{F}$ is called an *eigenvalue* of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

In the definition above, we require that $v \neq 0$ because every scalar $\lambda \in \mathbb{F}$ satisfies $T0 = \lambda 0$.

The word **eigenvalue** is half-German, half-English. The German prefix **eigen** means "own" in the sense of characterizing an intrinsic property.

The comments above show that V has a one-dimensional subspace invariant under T if and only if T has an eigenvalue.

Example 5.6 Eigenvalue

Define an operator $T \in \operatorname{End}(\mathbb{F}^3)$ by

$$T(x, y, z) = (7x + 3z,\ 3x + 6y + 9z,\ -6y)$$

for $(x, y, z) \in \mathbb{F}^3$. Then $T(3, 1, -1) = (18, 6, -6) = 6(3, 1, -1)$. Thus 6 is an eigenvalue of T.

Reminder: $I \in \operatorname{End}(V)$ is the identity operator. Thus $Iv = v$ for all $v \in V$.

The equivalences in the next result, along with many deep results in linear algebra, are valid only in the context of finite-dimensional vector spaces.

Equivalent conditions to be an eigenvalue — Result 5.7

Suppose V is finite-dimensional, $T \in \operatorname{End}(V)$, and $\lambda \in \mathbb{F}$. Then the following are equivalent.

(a) λ is an eigenvalue of T.

(b) $T - \lambda I$ is not injective.

Result 5.7 *continued*

(c) $T - \lambda I$ is not surjective.

(d) $T - \lambda I$ is not invertible.

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 3.65. □

Definition 5.8

Eigenvector

Suppose $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$ is an eigenvalue of T. A vector $v \in V$ is called an *eigenvector* of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

In other words, a nonzero vector $v \in V$ is an eigenvector of an operator $T \in \text{End}(V)$ if and only if Tv is a scalar multiple of 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 $v \in \ker(T - \lambda I)$.

Example 5.9 **Eigenvalues and eigenvectors**

Suppose $T \in \text{End}(\mathbb{F}^2)$ is defined by $T(w, z) = (-z, w)$.

(a) First consider the case $\mathbb{F} = \mathbb{R}$. Then T is a counterclockwise rotation by 90° about the origin in $\mathbb{R}^2$. 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 $\mathbb{R}^2$ cannot equal a scalar multiple of itself. Conclusion: if $\mathbb{F} = \mathbb{R}$, then T has no eigenvalues (and thus has no eigenvectors).

(b) Now consider the case $\mathbb{F} = \mathbb{C}$. 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 $T(w, z) = \lambda(w, z)$ is equivalent to the simultaneous equations

$$-z = \lambda w, \qquad w = \lambda z. \tag{5.10}$$

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 5.10 implies that $w = 0$; we are looking for solutions to 5.10 such that (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 can verify 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 $w \in \mathbb{C}$ and $w \neq 0$. Furthermore, the eigenvectors corresponding to the eigenvalue $-i$ are the vectors of the form (w, wi), with $w \in \mathbb{C}$ and $w \neq 0$.

In the next proof, we again use the equivalence

$$Tv = \lambda v \iff (T - \lambda I)v = 0.$$

Linearly independent eigenvectors — **Result 5.11**

Suppose $T \in \text{End}(V)$. Then every list of eigenvectors of T corresponding to distinct eigenvalues of T is linearly independent.

Proof. Suppose the desired result is false. Then there exists a smallest positive integer m such that there exists a linearly dependent list $v_1, \dots, v_m$ of eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_m$ of T (note that $m \geq 2$ because an eigenvector is, by definition, nonzero). Thus there exist $a_1, \dots, a_m \in \mathbb{F}$, none of which are 0 (because of the minimality of m), such that

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

Apply $T - \lambda_m I$ to both sides of the equation above, getting

$$a_1(\lambda_1 - \lambda_m)v_1 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

Because the eigenvalues $\lambda_1, \dots, \lambda_m$ are distinct, none of the coefficients above equal 0. Thus $v_1, \dots, v_{m-1}$ is a linearly dependent list of $m - 1$ eigenvectors of T corresponding to distinct eigenvalues, contradicting the minimality of m. This contradiction completes the proof. □

The result above leads to a short proof of the result below, which puts an upper bound on the number of distinct eigenvalues that an operator can have.

Operator cannot have more eigenvalues than dimension of vector space — **Result 5.12**

Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Proof. Let $T \in \text{End}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T. Let $v_1, \dots, v_m$ be corresponding eigenvectors. Then 5.11 implies that the list $v_1, \dots, v_m$ is linearly independent. Thus $m \leq \dim V$ (see 2.22), as desired. □

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. In this subsection we define that notion and the concept of applying a polynomial to an operator. This concept will be the key tool that we use in the next section when we prove that every operator on a nonzero finite-dimensional complex vector space has an eigenvalue.

If T is an operator, then TT makes sense (see 3.7) and is also an operator on the same vector space as T. We usually write T^2 instead of TT. More generally, we have the following definition of T^m.

Notation 5.13 T^m

Suppose $T \in \text{End}(V)$ and m is a positive integer.

- $T^m \in \text{End}(V)$ 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 $T^{-m} \in \text{End}(V)$ 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 arbitrary integers if T is invertible and are nonnegative integers if T is not invertible.

Having defined powers of an operator, we can now define what it means to apply a polynomial to an operator.

Notation 5.14 $p(T)$

Suppose $T \in \text{End}(V)$ and $p \in \mathbb{F}[X]$ is a polynomial given by

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

for all $z \in \mathbb{F}$. Then $p(T)$ is the operator on V 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 p to operators, not just elements of $\mathbb{F}$. The idea here is that to evaluate $p(T)$, we simply replace z with T in the expression defining p. Note that the constant term a_0 in $p(z)$ becomes the operator $a_0 I$ (which is a reasonable choice because $a_0 = a_0 z^0$ and thus we should replace a_0 with $a_0 T^0$, which equals $a_0 I$).

Example 5.15 **A polynomial applied to the differentiation operator**

Suppose $D \in \text{End}(\mathbb{R}[X])$ 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 $q \in \mathbb{R}[X]$.

If we fix an operator $T \in \text{End}(V)$, then the function from $\mathbb{F}[X]$ to $\text{End}(V)$ given by $p \mapsto p(T)$ is linear, as you should verify.

Product of polynomials
If $p, q \in \mathbb{F}[X]$, then $pq \in \mathbb{F}[X]$ is the polynomial defined by

$$(pq)(z) = p(z)q(z)$$

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

Definition 5.16

Informal proof: When a product of polynomials is expanded using the distributive property, it does not matter whether the symbol is z or T.

The order does not matter in taking products of polynomials of a single operator, as shown by (b) in the next result.

Multiplicative properties
Suppose $p, q \in \mathbb{F}[X]$ and $T \in \text{End}(V)$. Then

(a) $(pq)(T) = p(T)q(T)$;

(b) $p(T)q(T) = q(T)p(T)$.

Result 5.17

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 all $z \in \mathbb{F}$. 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) Using (a) twice, we have $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$. □

We observed earlier that if $T \in \text{End}(V)$, then the subspaces $\ker T$ and $\text{im}\, T$ are invariant under T (see 5.4). Now we show that the null space and the range of every polynomial of T are also invariant under T.

Null space and range of $p(T)$ are invariant under T
Suppose $T \in \text{End}(V)$ and $p \in \mathbb{F}[X]$. Then $\ker p(T)$ and $\text{im}\, p(T)$ are invariant under T.

Result 5.18

Proof. Suppose $u \in \ker p(T)$. Then $p(T)u = 0$. Thus

$$\big(p(T)\big)(Tu) = \big(p(T)\, T\big)(u) = \big(T\, p(T)\big)(u) = T\big(p(T)u\big) = T(0) = 0.$$

Hence $Tu \in \ker p(T)$. Thus $\ker p(T)$ is invariant under T, as desired.

Suppose $u \in \operatorname{im} p(T)$. Then there exists $v \in V$ such that $u = p(T)v$. Thus

$$Tu = T\big(p(T)v\big) = p(T)(Tv).$$

Hence $Tu \in \operatorname{im} p(T)$. Thus $\operatorname{im} p(T)$ is invariant under T, as desired. □

Exercises

5A.1 Suppose $T \in \operatorname{End}(V)$ and U is a subspace of V.
(a) Prove that if $U \subseteq \ker T$, then U is invariant under T.
(b) Prove that if $\operatorname{im} T \subseteq U$, then U is invariant under T.

5A.2 Suppose that $T \in \operatorname{End}(V)$ and $V_1, \ldots, V_m$ are subspaces of V invariant under T. Prove that $V_1 + \cdots + V_m$ is invariant under T.

5A.3 Suppose $T \in \operatorname{End}(V)$. Prove that the intersection of every collection of subspaces of V invariant under T is invariant under T.

5A.4 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 $U = V$.

5A.5 Suppose $T \in \operatorname{End}(\mathbb{R}^2)$ is defined by $T(x, y) = (-3y, x)$. Find the eigenvalues of T.

5A.6 Define $T \in \operatorname{End}(\mathbb{F}^2)$ by $T(w, z) = (z, w)$. Find all eigenvalues and eigenvectors of T.

5A.7 Define $T \in \operatorname{End}(\mathbb{F}^3)$ by $T(z_1, z_2, z_3) = (2z_2, 0, 5z_3)$. Find all eigenvalues and eigenvectors of T.

5A.8 Suppose $P \in \operatorname{End}(V)$ is such that $P^2 = P$. Prove that if λ is an eigenvalue of P, then $\lambda = 0$ or $\lambda = 1$.

5A.9 Define $T\colon \mathbb{R}[X] \to \mathbb{R}[X]$ by $Tp = p'$. Find all eigenvalues and eigenvectors of T.

5A.10 Define $T \in \operatorname{End}(\mathbb{R}[X]_{\leq 4})$ by $(Tp)(x) = xp'(x)$ for all $x \in \mathbb{R}$. Find all eigenvalues and eigenvectors of T.

5A.11 Suppose V is finite-dimensional, $T \in \operatorname{End}(V)$, and $\alpha \in \mathbb{F}$. Prove that there exists $\delta > 0$ such that $T - \lambda I$ is invertible for all $\lambda \in \mathbb{F}$ such that $0 < |\alpha - \lambda| < \delta$.

5A.12 Suppose $V = U \oplus W$, where U and W are nonzero subspaces of V. Define $P \in \text{End}(V)$ by $P(u + w) = u$ for each $u \in U$ and each $w \in W$. Find all eigenvalues and eigenvectors of P.

5A.13 Suppose $T \in \text{End}(V)$. Suppose $S \in \text{End}(V)$ 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 $S^{-1}TS$?

5A.14 Give an example of an operator on $\mathbb{R}^4$ that has no (real) eigenvalues.

5A.15 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and $\lambda \in \mathbb{F}$. Show that λ is an eigenvalue of T if and only if λ is an eigenvalue of the dual operator $T' \in \text{End}(V')$.

5A.16 Suppose $v_1, \dots, v_n$ is a basis of V and $T \in \text{End}(V)$. Prove that if λ is an eigenvalue of T, then

$$|\lambda| \le n \max\{|\text{Mat}(T)_{j,k}| : 1 \le j, k \le n\},$$

where $\text{Mat}(T)_{j,k}$ denotes the entry in row j, column k of the matrix of T with respect to the basis $v_1, \dots, v_n$.

See Exercise 19 in Section 6A for a different bound on $|\lambda|$.

5A.17 Suppose $\mathbb{F} = \mathbb{R}$, $T \in \text{End}(V)$, and $\lambda \in \mathbb{R}$. Prove that λ is an eigenvalue of T if and only if λ is an eigenvalue of the complexification $T_{\mathbb{C}}$.

See Exercise 33 in Section 3B for the definition of $T_{\mathbb{C}}$.

5A.18 Suppose $\mathbb{F} = \mathbb{R}$, $T \in \text{End}(V)$, and $\lambda \in \mathbb{C}$. Prove that λ is an eigenvalue of the complexification $T_{\mathbb{C}}$ if and only if $\overline{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$.

5A.19 Show that the forward shift operator $T \in \text{End}(\mathbb{F}^\infty)$ defined by

$$T(z_1, z_2, \dots) = (0, z_1, z_2, \dots)$$

has no eigenvalues.

5A.20 Define the backward shift operator $S \in \text{End}(\mathbb{F}^\infty)$ by

$$S(z_1, z_2, z_3, \dots) = (z_2, z_3, \dots).$$

(a) Show that every element of $\mathbb{F}$ is an eigenvalue of S.

(b) Find all eigenvectors of S.

5A.21 Suppose $T \in \text{End}(V)$ is invertible.

(a) Suppose $\lambda \in \mathbb{F}$ with $\lambda \neq 0$. Prove that λ is an eigenvalue of T if and only if $1/\lambda$ is an eigenvalue of T^{-1}.

(b) Prove that T and T^{-1} have the same eigenvectors.

5A.22 Suppose $T \in \text{End}(V)$ and there exist nonzero vectors u and w in V such that

$$Tu = 3w \quad \text{and} \quad Tw = 3u.$$

Prove that 3 or -3 is an eigenvalue of T.

5A.23 Suppose V is finite-dimensional and $S, T \in \text{End}(V)$. Prove that ST and TS have the same eigenvalues.

5A.24 Suppose A is an n-by-n matrix with entries in $\mathbb{F}$. Define $T \in \text{End}(\mathbb{F}^n)$ by $Tx = Ax$, where elements of $\mathbb{F}^n$ 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.

5A.25 Suppose $T \in \text{End}(V)$ and u, w are eigenvectors of T such that $u + w$ is also an eigenvector of T. Prove that u and w are eigenvectors of T corresponding to the same eigenvalue.

5A.26 Suppose $T \in \text{End}(V)$ is such that every nonzero vector in V is an eigenvector of T. Prove that T is a scalar multiple of the identity operator.

5A.27 Suppose that V is finite-dimensional and $k \in \{1, \ldots, \dim V - 1\}$. Suppose $T \in \text{End}(V)$ is such that every subspace of V of dimension k is invariant under T. Prove that T is a scalar multiple of the identity operator.

5A.28 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that T has at most $1 + \dim \text{im}\, T$ distinct eigenvalues.

5A.29 Suppose $T \in \text{End}(\mathbb{R}^3)$ and -4, 5, and $\sqrt{7}$ are eigenvalues of T. Prove that there exists $x \in \mathbb{R}^3$ such that $Tx - 9x = (-4, 5, \sqrt{7})$.

5A.30 Suppose $T \in \text{End}(V)$ and $(T - 2I)(T - 3I)(T - 4I) = 0$. Suppose λ is an eigenvalue of T. Prove that $\lambda = 2$ or $\lambda = 3$ or $\lambda = 4$.

5A.31 Give an example of $T \in \text{End}(\mathbb{R}^2)$ such that $T^4 = -I$.

5A.32 Suppose $T \in \text{End}(V)$ has no eigenvalues and $T^4 = I$. Prove that $T^2 = -I$.

5A.33 Suppose $T \in \text{End}(V)$ and m is a positive integer.
(a) Prove that T is injective if and only if T^m is injective.
(b) Prove that T is surjective if and only if T^m is surjective.

5A.34 Suppose V is finite-dimensional and $v_1, \dots, v_m \in V$. Prove that the list $v_1, \dots, v_m$ is linearly independent if and only if there exists $T \in \text{End}(V)$ such that $v_1, \dots, v_m$ are eigenvectors of T corresponding to distinct eigenvalues.

5A.35 Suppose that $\lambda_1, \dots, \lambda_n$ is a list of distinct real numbers. Prove that the list $e^{\lambda_1 x}, \dots, e^{\lambda_n x}$ is linearly independent in the vector space of real-valued functions on $\mathbb{R}$.

Hint: Let $V = \text{span}(e^{\lambda_1 x}, \dots, e^{\lambda_n x})$, and define an operator $D \in \text{End}(V)$ by $Df = f'$. Find eigenvalues and eigenvectors of D.

5A.36 Suppose that $\lambda_1, \dots, \lambda_n$ is a list of distinct positive numbers. Prove that the list $\cos(\lambda_1 x), \dots, \cos(\lambda_n x)$ is linearly independent in the vector space of real-valued functions on $\mathbb{R}$.

5A.37 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Define $\mathcal{A} \in \text{End}(\text{End}(V))$ by

$$\mathcal{A}(S) = TS$$

for each $S \in \text{End}(V)$. Prove that the set of eigenvalues of T equals the set of eigenvalues of $\mathcal{A}$.

5A.38 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and U is a subspace of V invariant under T. The *quotient operator* $T/U \in \text{End}(V/U)$ is defined by

$$(T/U)(v + U) = Tv + U$$

for each $v \in V$.
(a) Show that the definition of T/U makes sense (which requires using the condition that U is invariant under T) and show that T/U is an operator on V/U.
(b) Show that each eigenvalue of T/U is an eigenvalue of T.

5A.39 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that T has an eigenvalue if and only if there exists a subspace of V of dimension $\dim V - 1$ that is invariant under T.

5A.40 Suppose $S, T \in \text{End}(V)$ and S is invertible. Suppose $p \in \mathbb{F}[X]$ is a polynomial. Prove that

$$p(STS^{-1}) = Sp(T)S^{-1}.$$

5A.41 Suppose $T \in \text{End}(V)$ and U is a subspace of V invariant under T. Prove that U is invariant under $p(T)$ for every polynomial $p \in \mathbb{F}[X]$.

5A.42 Define $T \in \text{End}(\mathbb{F}^n)$ by $T(x_1, x_2, x_3, \ldots, x_n) = (x_1, 2x_2, 3x_3, \ldots, nx_n)$.
(a) Find all eigenvalues and eigenvectors of T.
(b) Find all subspaces of $\mathbb{F}^n$ that are invariant under T.

5A.43 Suppose that V is finite-dimensional, $\dim V > 1$, and $T \in \text{End}(V)$. Prove that $\{p(T) : p \in \mathbb{F}[X]\} \neq \text{End}(V)$.

5B The Minimal Polynomial

Existence of Eigenvalues on Complex Vector Spaces

Now we come to one of the central results about operators on finite-dimensional complex vector spaces.

Existence of eigenvalues **Result 5.19**
Every operator on a finite-dimensional nonzero complex vector space has an eigenvalue.

Proof. Suppose V is a finite-dimensional complex vector space of dimension $n > 0$ and $T \in \text{End}(V)$. 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 this list has length $n + 1$. Hence some linear combination (with not all the coefficients equal to 0) of the vectors above equals 0. Thus there exists a nonconstant polynomial p of smallest degree such that

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

By the first version of the fundamental theorem of algebra (see 4.12), there exists $\lambda \in \mathbb{C}$ such that $p(\lambda) = 0$. Hence there exists a polynomial $q \in \mathbb{C}[X]$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbb{C}$ (see 4.6). This implies (using 5.17) that

$$0 = p(T)v = (T - \lambda I)\big(q(T)v\big).$$

Because q has smaller degree than p, we know that $q(T)v \neq 0$. Thus the equation above implies that λ is an eigenvalue of T with eigenvector $q(T)v$. □

The proof above makes crucial use of the fundamental theorem of algebra. The comment following Exercise 16 helps explain why the fundamental theorem of algebra is so tightly connected to the result above.

The hypothesis in the result above that $\mathbb{F} = \mathbb{C}$ cannot be replaced with the hypothesis that $\mathbb{F} = \mathbb{R}$, as shown by Example 5.9. The next example shows that the finite-dimensional hypothesis in the result above also cannot be deleted.

Example 5.20 **An operator on a complex vector space with no eigenvalues**

Define $T \in \text{End}(\mathbb{C}[X])$ by $(Tp)(z) = zp(z)$. If $p \in \mathbb{C}[X]$ is a nonzero polynomial, then the degree of Tp is one more than the degree of p, and thus Tp cannot equal a scalar multiple of p. Hence T has no eigenvalues.

Because $\mathbb{C}[X]$ is infinite-dimensional, this example does not contradict the result above.

Eigenvalues and the Minimal Polynomial

In this subsection we introduce an important polynomial associated with each operator. We begin with the following definition.

Monic polynomial — Definition 5.21

A *monic polynomial* is a polynomial whose highest-degree coefficient equals 1.

For example, the polynomial $2 + 9z^2 + z^7$ is a monic polynomial of degree 7.

Existence, uniqueness, and degree of minimal polynomial — Result 5.22

Suppose V is finite-dimensional and $T \in \text{End}(V)$. Then there is a unique monic polynomial $p \in \mathbb{F}[X]$ of smallest degree such that $p(T) = 0$. Furthermore, $\deg p \le \dim V$.

Proof. If $\dim V = 0$, then I is the zero operator on V and thus we take p to be the constant polynomial 1.

Now use induction on $\dim V$. Thus assume that $\dim V > 0$ and that the desired result is true for all operators on all vector spaces of smaller dimension. Let $u \in V$ be such that $u \ne 0$. The list $u, Tu, \dots, T^{\dim V}u$ has length $1 + \dim V$ and thus is linearly dependent. By the linear dependence lemma (2.19), there is a smallest positive integer $m \le \dim V$ such that $T^m u$ is a linear combination of $u, Tu, \dots, T^{m-1}u$. Thus there exist scalars $c_0, c_1, c_2, \dots, c_{m-1} \in \mathbb{F}$ such that

$$c_0 u + c_1 Tu + \cdots + c_{m-1}T^{m-1}u + T^m u = 0. \tag{5.23}$$

Define a monic polynomial $q \in \mathbb{F}[X]_{\le m}$ by

$$q(z) = c_0 + c_1 z + \cdots + c_{m-1}z^{m-1} + z^m.$$

Then 5.23 implies that $q(T)u = 0$.

If k is a nonnegative integer, then

$$q(T)(T^k u) = T^k\big(q(T)u\big) = T^k(0) = 0.$$

The linear dependence lemma (2.19) shows that $u, Tu, \dots, T^{m-1}u$ is linearly independent. Thus the equation above implies that $\dim \ker q(T) \geq m$. Hence

$$\dim \operatorname{im} q(T) = \dim V - \dim \ker q(T) \leq \dim V - m.$$

Because $\operatorname{im} q(T)$ is invariant under T (by 5.18), we can apply our induction hypothesis to the operator $T|_{\operatorname{im} q(T)}$ on the vector space $\operatorname{im} q(T)$. Thus there is a monic polynomial $s \in \mathbb{F}[X]$ with

$$\deg s \leq \dim V - m \quad \text{and} \quad s(T|_{\operatorname{im} q(T)}) = 0.$$

Hence for all $v \in V$ we have

$$\big((sq)(T)\big)(v) = s(T)\big(q(T)v\big) = 0$$

because $q(T)v \in \operatorname{im} q(T)$ and $s(T)|_{\operatorname{im} q(T)} = s(T|_{\operatorname{im} q(T)}) = 0$. Thus sq is a monic polynomial such that $\deg sq \leq \dim V$ and $(sq)(T) = 0$.

The paragraph above shows that there is a monic polynomial of degree at most $\dim V$ that when applied to T gives the 0 operator. Thus there is a monic polynomial of smallest degree with this property, completing the existence part of this result.

Let $p \in \mathbb{F}[X]$ be a monic polynomial of smallest degree such that $p(T) = 0$. To prove the uniqueness part of the result, suppose $r \in \mathbb{F}[X]$ is a monic polynomial of the same degree as p and $r(T) = 0$. Then $(p-r)(T) = 0$ and also $\deg(p-r) < \deg p$. If $p - r$ were not equal to 0, then we could divide $p - r$ by the coefficient of the highest-order term in $p - r$ to get a monic polynomial (of smaller degree than p) that when applied to T gives the 0 operator. Thus $p - r = 0$, as desired. □

The previous result justifies the following definition.

Minimal polynomial — Definition 5.24

Suppose V is finite-dimensional and $T \in \operatorname{End}(V)$. Then the *minimal polynomial* of T is the unique monic polynomial $p \in \mathbb{F}[X]$ of smallest degree such that $p(T) = 0$.

To compute the minimal polynomial of an operator $T \in \operatorname{End}(V)$, we need to find the smallest positive integer m such that the equation

$$c_0 I + c_1 T + \cdots + c_{m-1} T^{m-1} = -T^m$$

has a solution $c_0, c_1, \dots, c_{m-1} \in \mathbb{F}$. If we pick a basis of V and replace T in the equation above with the matrix of T, then the equation above can be thought of as a system of $(\dim V)^2$ linear equations in the m unknowns $c_0, c_1, \dots, c_{m-1} \in \mathbb{F}$. Gaussian elimination or another fast method of solving systems of linear equations can tell us whether a solution exists, testing successive values $m = 1, 2, \dots$ until a solution exists.

By 5.22, a solution exists for some smallest positive integer $m \le \dim V$. The minimal polynomial of T is then $c_0 + c_1 z + \cdots + c_{m-1} z^{m-1} + z^m$.

Even faster (usually), pick $v \in V$ with $v \ne 0$ and consider the equation

$$c_0 v + c_1 T v + \cdots + c_{\dim V - 1} T^{\dim V - 1} v = -T^{\dim V} v. \tag{5.25}$$

Use a basis of V to convert the equation above to a system of $\dim V$ linear equations in $\dim V$ unknowns $c_0, c_1, \ldots, c_{\dim V - 1}$. If this system of equations has a unique solution $c_0, c_1, \ldots, c_{\dim V - 1}$ (as happens most of the time), then the scalars $c_0, c_1, \ldots, c_{\dim V - 1}, 1$ are the coefficients of the minimal polynomial of T (because 5.22 states that the degree of the minimal polynomial is at most $\dim V$).

Consider operators on $\mathbb{R}^4$ (thought of as 4-by-4 matrices with respect to the standard basis), and take $v = (1, 0, 0, 0)$ in the paragraph above. The faster method described above works on over 99.8% of the 4-by-4 matrices with integer entries in the interval $[-10, 10]$ and on over 99.999% of the 4-by-4 matrices with integer entries in $[-100, 100]$.

These estimates are based on testing millions of random matrices.

The next example illustrates the faster procedure discussed above.

Example 5.26 **Minimal polynomial of an operator on $\mathbb{F}^5$**

Suppose $T \in \text{End}(\mathbb{F}^5)$ and

$$\text{Mat}(T) = \begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

with respect to the standard basis e_1, e_2, e_3, e_4, e_5. Taking $v = e_1$ for 5.25, we have

$$\begin{aligned} Te_1 &= e_2, & T^4 e_1 &= T(T^3 e_1) = Te_4 = e_5, \\ T^2 e_1 &= T(Te_1) = Te_2 = e_3, & T^5 e_1 &= T(T^4 e_1) = Te_5 = -3e_1 + 6e_2. \\ T^3 e_1 &= T(T^2 e_1) = Te_3 = e_4, \end{aligned}$$

Thus $3e_1 - 6Te_1 = -T^5 e_1$. The list $e_1, Te_1, T^2 e_1, T^3 e_1, T^4 e_1$, which equals the list e_1, e_2, e_3, e_4, e_5, is linearly independent, so no other linear combination of this list equals $-T^5 e_1$. Hence the minimal polynomial of T is $3 - 6z + z^5$.

Recall that by definition, eigenvalues of operators on V and zeros of polynomials in $\mathbb{F}[X]$ must be elements of $\mathbb{F}$. In particular, if $\mathbb{F} = \mathbb{R}$, then eigenvalues and zeros must be real numbers.

Result 5.27 **Eigenvalues are the zeros of the minimal polynomial**

Suppose V is finite-dimensional and $T \in \text{End}(V)$.

(a) The zeros of the minimal polynomial of T are the eigenvalues of T.

Result 5.27 *continued*

(b) If V is a complex vector space, then the minimal polynomial of T has the form

$$(z-\lambda_1)\cdots(z-\lambda_m),$$

where $\lambda_1,\ldots,\lambda_m$ is a list of all eigenvalues of T, possibly with repetitions.

Proof. Let p be the minimal polynomial of T.

(a) First suppose $\lambda \in \mathbb{F}$ 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 $\mathbb{F}$ (see 4.6). Because $p(T) = 0$, we have

$$0 = (T-\lambda I)\big(q(T)v\big)$$

for all $v \in V$. Because $\deg q = (\deg p) - 1$ and p is the minimal polynomial of T, 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 that every eigenvalue of T is a zero of p, now suppose $\lambda \in \mathbb{F}$ 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 $T^k v = \lambda^k v$ for every nonnegative integer k. Thus

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

Because p is the minimal polynomial of T, we have $p(T)v = 0$. Hence the equation above implies that $p(\lambda) = 0$. Thus λ is a zero of p, as desired.

(b) To get the desired result, use (a) and the second version of the fundamental theorem of algebra (see 4.13). □

A nonzero polynomial has at most as many distinct zeros as its degree (see 4.8). Thus (a) of the previous result, along with the result that the minimal polynomial of an operator on V has degree at most $\dim V$, gives an alternative proof of 5.12, which states that an operator on V has at most $\dim V$ distinct eigenvalues.

Every monic polynomial is the minimal polynomial of some operator, as shown by Exercise 16, which generalizes Example 5.26. Thus 5.27(a) shows that finding exact expressions for the eigenvalues of an operator is equivalent to the problem of finding exact expressions for the zeros of a polynomial (and thus is not possible for some operators).

Example 5.28 **An operator whose eigenvalues cannot be found exactly**

Let $T \in \text{End}(\mathbb{C}^5)$ be the operator defined by

$$T(z_1,z_2,z_3,z_4,z_5) = (-3z_5,\, z_1+6z_5,\, z_2,\, z_3,\, z_4).$$

The matrix of T with respect to the standard basis of $\mathbb{C}^5$ is the 5-by-5 matrix in Example 5.26. As we showed in that example, the minimal polynomial of T is

Example 5.28
continued

the polynomial

$$3 - 6z + z^5.$$

No zero of the polynomial above can be expressed using rational numbers, roots of rational numbers, and the usual rules of arithmetic (a proof of this would take us considerably beyond linear algebra). Because the zeros of the polynomial above are the eigenvalues of T [by 5.27(a)], we cannot find an exact expression for any eigenvalue of T in any familiar form.

Numerical techniques, which we will not discuss here, show that the zeros of the polynomial above, and thus the eigenvalues of T, are approximately the following five complex numbers:

$$-1.67, \quad 0.51, \quad 1.40, \quad -0.12 + 1.59i, \quad -0.12 - 1.59i.$$

Note that the two nonreal zeros of this polynomial are complex conjugates of each other, as we expect for a polynomial with real coefficients (see 4.14).

The next result completely characterizes the polynomials that when applied to an operator give the 0 operator.

$q(T) = 0 \iff q$ is a polynomial multiple of the minimal polynomial Result 5.29

Suppose V is finite-dimensional, $T \in \text{End}(V)$, and $q \in \mathbb{F}[X]$. 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 suppose $q(T) = 0$. By the division algorithm for polynomials (4.9), there exist polynomials $s, r \in \mathbb{F}[X]$ such that

$$q = ps + r \tag{5.30}$$

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 5.30 becomes the equation $q = ps$. Hence q is a polynomial multiple of p, as desired.

To prove the other direction, now suppose q is a polynomial multiple of p. Thus there exists a polynomial $s \in \mathbb{F}[X]$ such that $q = ps$. We have

$$q(T) = p(T)s(T) = 0\, s(T) = 0,$$

as desired. □

The next result is a nice consequence of the result above.

Minimal polynomial of a restriction operator — Result 5.31

Suppose V is finite-dimensional, $T \in \text{End}(V)$, and U is a subspace of V that is invariant under T. Then the minimal polynomial of T is a polynomial multiple of the minimal polynomial of $T|_U$.

Proof. Suppose p is the minimal polynomial of T. Thus $p(T)v = 0$ for all $v \in V$. In particular,

$$p(T)u = 0 \text{ for all } u \in U.$$

Thus $p(T|_U) = 0$. Now 5.29, applied to the operator $T|_U$ in place of T, implies that p is a polynomial multiple of the minimal polynomial of $T|_U$. □

See Exercise 25 for a result about quotient operators that is analogous to the result above.

The next result shows that the constant term of the minimal polynomial of an operator determines whether the operator is invertible.

T not invertible $\iff$ constant term of minimal polynomial of T is 0 — Result 5.32

Suppose V is finite-dimensional and $T \in \text{End}(V)$. Then T is not invertible if and only if the constant term of the minimal polynomial of T is 0.

Proof. Suppose $T \in \text{End}(V)$ and p is the minimal polynomial of T. Then

$$\begin{aligned} T \text{ is not invertible} &\iff 0 \text{ is an eigenvalue of } T \\ &\iff 0 \text{ is a zero of } p \\ &\iff \text{the constant term of } p \text{ is } 0, \end{aligned}$$

where the first equivalence holds by 5.7, the second equivalence holds by 5.27(a), and the last equivalence holds because the constant term of p equals $p(0)$. □

Eigenvalues on Odd-Dimensional Real Vector Spaces

The next result will be the key tool that we use to show that every operator on an odd-dimensional real vector space has an eigenvalue.

Even-dimensional null space — Result 5.33

Suppose $\mathbb{F} = \mathbb{R}$ and V is finite-dimensional. Suppose also that $T \in \text{End}(V)$ and $b, c \in \mathbb{R}$ with $b^2 < 4c$. Then $\dim \ker(T^2 + bT + cI)$ is an even number.

Proof. Recall that $\ker(T^2 + bT + cI)$ is invariant under T (by 5.18). By replacing V with $\ker(T^2 + bT + cI)$ and replacing T with T restricted to $\ker(T^2 + bT + cI)$, we can assume that $T^2 + bT + cI = 0$; we now need to prove that $\dim V$ is even.

Suppose $\lambda \in \mathbb{R}$ and $v \in V$ are such that $Tv = \lambda v$. Then

$$0 = (T^2 + bT + cI)v = (\lambda^2 + b\lambda + c)v = \left((\lambda + \tfrac{b}{2})^2 + c - \tfrac{b^2}{4}\right)v.$$

The term in large parentheses above is a positive number. Thus the equation above implies that $v = 0$. Hence we have shown that T has no eigenvectors.

Let U be a subspace of V that is invariant under T and has the largest dimension among all subspaces of V that are invariant under T and have even dimension. If $U = V$, then we are done; otherwise assume there exists $w \in V$ such that $w \notin U$.

Let $W = \operatorname{span}(w, Tw)$. Then W is invariant under T because $T(Tw) = -bTw - cw$. Furthermore, $\dim W = 2$ because otherwise w would be an eigenvector of T. Now

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) = \dim U + 2,$$

where $U \cap W = \{0\}$ because otherwise $U \cap W$ would be a one-dimensional subspace of V that is invariant under T (impossible because T has no eigenvectors).

Because $U + W$ is invariant under T, the equation above shows that there exists a subspace of V invariant under T of even dimension larger than $\dim U$. Thus the assumption that $U \neq V$ was incorrect. Hence V has even dimension. □

The next result states that on odd-dimensional vector spaces, every operator has an eigenvalue. We already know this result for finite-dimensional complex vector spaces (without the odd hypothesis). Thus in the proof below, we will assume that $\mathbb{F} = \mathbb{R}$.

Operators on odd-dimensional vector spaces have eigenvalues **Result 5.34**

Every operator on an odd-dimensional vector space has an eigenvalue.

Proof. Suppose $\mathbb{F} = \mathbb{R}$ and V is finite-dimensional. Let $n = \dim V$, and suppose n is an odd number. Let $T \in \operatorname{End}(V)$. We will use induction on n in steps of size two to show that T has an eigenvalue. To get started, note that the desired result holds if $\dim V = 1$ because then every nonzero vector in V is an eigenvector of T.

Now suppose that $n \geq 3$ and the desired result holds for all operators on all odd-dimensional vector spaces of dimension less than n. Let p denote the minimal polynomial of T. If p is a polynomial multiple of $x - \lambda$ for some $\lambda \in \mathbb{R}$, then λ is an eigenvalue of T [by 5.27(a)] and we are done. Thus we can assume that there exist $b, c \in \mathbb{R}$ such that $b^2 < 4c$ and p is a polynomial multiple of $x^2 + bx + c$ (see 4.16).

There exists a monic polynomial $q \in \mathbb{R}[X]$ such that $p(x) = q(x)(x^2 + bx + c)$ for all $x \in \mathbb{R}$. Now

$$0 = p(T) = \big(q(T)\big)(T^2 + bT + cI),$$

which means that $q(T)$ equals 0 on $\operatorname{im}(T^2 + bT + cI)$. Because $\deg q < \deg p$ and p is the minimal polynomial of T, this implies that $\operatorname{im}(T^2 + bT + cI) \neq V$.

The fundamental theorem of linear maps (3.21) tells us that

$$\dim V = \dim \ker(T^2 + bT + cI) + \dim \operatorname{im}(T^2 + bT + cI).$$

Because $\dim V$ is odd (by hypothesis) and $\dim \ker(T^2 + bT + cI)$ is even (by 5.33), the equation above shows that $\dim \operatorname{im}(T^2 + bT + cI)$ is odd.

Hence $\operatorname{im}(T^2 + bT + cI)$ is a subspace of V that is invariant under T (by 5.18) and that has odd dimension less than $\dim V$. Our induction hypothesis now implies that T restricted to $\operatorname{im}(T^2 + bT + cI)$ has an eigenvalue, which means that T has an eigenvalue. □

See Exercise 23 in Section 8B and Exercise 10 in Section 9C for alternative proofs of the result above.

Exercises

5B.1 Suppose $T \in \text{End}(V)$. Prove that 9 is an eigenvalue of T^2 if and only if 3 or -3 is an eigenvalue of T.

5B.2 Suppose V is a complex vector space and $T \in \text{End}(V)$ has no eigenvalues. Prove that every subspace of V invariant under T is either $\{0\}$ or infinite-dimensional.

5B.3 Suppose n is an integer with $n > 1$ and $T \in \text{End}(\mathbb{F}^n)$ is defined by

$$T(x_1, \dots, x_n) = (x_1 + \cdots + x_n, \dots, x_1 + \cdots + x_n).$$

(a) Find all eigenvalues and eigenvectors of T.

(b) Find the minimal polynomial of T.

The matrix of T with respect to the standard basis of $\mathbb{F}^n$ consists of all 1's.

5B.4 Suppose $\mathbb{F} = \mathbb{C}$, $T \in \text{End}(V)$, $p \in \mathbb{C}[X]$ is a nonconstant polynomial, and $\alpha \in \mathbb{C}$. Prove that α is an eigenvalue of $p(T)$ if and only if $\alpha = p(\lambda)$ for some eigenvalue λ of T.

5B.5 Give an example of an operator on $\mathbb{R}^2$ that shows the result in Exercise 4 does not hold if $\mathbb{C}$ is replaced with $\mathbb{R}$.

5B.6 Suppose $T \in \text{End}(\mathbb{F}^2)$ is defined by $T(w, z) = (-z, w)$. Find the minimal polynomial of T.

5B.7

(a) Give an example of $S, T \in \text{End}(\mathbb{F}^2)$ such that the minimal polynomial of ST does not equal the minimal polynomial of TS.

(b) Suppose V is finite-dimensional and $S, T \in \text{End}(V)$. Prove that if at least one of S, T is invertible, then the minimal polynomial of ST equals the minimal polynomial of TS.

Hint: Show that if S is invertible and $p \in \mathbb{F}[X]$, then $p(TS) = S^{-1} p(ST) S$.

5B.8 Suppose $T \in \text{End}(\mathbb{R}^2)$ is the operator of counterclockwise rotation by 1°. Find the minimal polynomial of T.

Because $\dim \mathbb{R}^2 = 2$, the degree of the minimal polynomial of T is at most 2. Thus the minimal polynomial of T is not the tempting polynomial $x^{180} + 1$, even though $T^{180} = -I$.

5B.9 Suppose $T \in \text{End}(V)$ is such that with respect to some basis of V, all entries of the matrix of T are rational numbers. Explain why all coefficients of the minimal polynomial of T are rational numbers.

5B.10 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and $v \in V$. Prove that

$$\text{span}(v, Tv, \ldots, T^m v) = \text{span}(v, Tv, \ldots, T^{\dim V - 1} v)$$

for all integers $m \geq \dim V - 1$.

5B.11 Suppose V is a two-dimensional vector space, $T \in \text{End}(V)$, and the matrix of T with respect to some basis of V is $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$.

(a) Show that $T^2 - (a+d)T + (ad - bc)I = 0$.

(b) Show that the minimal polynomial of T equals

$$\begin{cases} z - a & \text{if } b = c = 0 \text{ and } a = d, \\ z^2 - (a+d)z + (ad - bc) & \text{otherwise.} \end{cases}$$

5B.12 Define $T \in \text{End}(\mathbb{F}^n)$ by $T(x_1, x_2, x_3, \ldots, x_n) = (x_1, 2x_2, 3x_3, \ldots, nx_n)$. Find the minimal polynomial of T.

5B.13 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and $p \in \mathbb{F}[X]$. Prove that there exists a unique $r \in \mathbb{F}[X]$ such that $p(T) = r(T)$ and $\deg r$ is less than the degree of the minimal polynomial of T.

5B.14 Suppose V is finite-dimensional and $T \in \text{End}(V)$ has minimal polynomial $4 + 5z - 6z^2 - 7z^3 + 2z^4 + z^5$. Find the minimal polynomial of T^{-1}.

5B.15 Suppose V is a finite-dimensional complex vector space with $\dim V > 0$ and $T \in \text{End}(V)$. Define $f : \mathbb{C} \to \mathbb{R}$ by

$$f(\lambda) = \dim \text{im}(T - \lambda I).$$

Prove that f is not a continuous function.

5B.16 Suppose $a_0, \ldots, a_{n-1} \in \mathbb{F}$. Let T be the operator on $\mathbb{F}^n$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & -a_2 \\ & & \ddots & & \vdots \\ & & & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{pmatrix}.$$

Here all entries of the matrix are 0 except for all 1's on the line under the diagonal and the entries in the last column (some of which might also be 0). Show that the minimal polynomial of T is the polynomial

$$a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n.$$

The matrix above is called the ***companion matrix*** *of the polynomial above. This exercise shows that every monic polynomial is the minimal polynomial of some operator. Hence a formula or an algorithm that could produce exact eigenvalues for each operator on each $\mathbb{F}^n$ could then produce exact zeros for each polynomial [by 5.27(a)]. Thus there is no such formula or algorithm. However, efficient numerical methods exist for obtaining very good approximations for the eigenvalues of an operator.*

5B.17 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and p is the minimal polynomial of T. Suppose $\lambda \in \mathbb{F}$. Show that the minimal polynomial of $T - \lambda I$ is the polynomial q defined by $q(z) = p(z + \lambda)$.

5B.18 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and p is the minimal polynomial of T. Suppose $\lambda \in \mathbb{F} \setminus \{0\}$. Show that the minimal polynomial of λT is the polynomial q defined by $q(z) = \lambda^{\deg p}\, p\Big(z/\lambda\Big)$.

5B.19 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Let $\mathscr{E}$ be the subspace of $\text{End}(V)$ defined by

$$\mathscr{E} = \{q(T) : q \in \mathbb{F}[X]\}.$$

Prove that $\dim \mathscr{E}$ equals the degree of the minimal polynomial of T.

5B.20 Suppose $T \in \text{End}(\mathbb{F}^4)$ is such that the eigenvalues of T are $3, 5, 8$. Prove that $(T - 3I)^2(T - 5I)^2(T - 8I)^2 = 0$.

5B.21 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that the minimal polynomial of T has degree at most $1 + \dim \text{im}\, T$.

If $\dim \text{im}\, T < \dim V - 1$, then this exercise gives a better upper bound than 5.22 for the degree of the minimal polynomial of T.

5B.22 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that T is invertible if and only if $I \in \text{span}(T, T^2, \ldots, T^{\dim V})$.

5B.23 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Let $n = \dim V$. Prove that if $v \in V$, then $\text{span}(v, Tv, \ldots, T^{n-1}v)$ is invariant under T.

5B.24 Suppose V is a finite-dimensional complex vector space. Suppose $T \in \text{End}(V)$ is such that 5 and 6 are eigenvalues of T and that T has no other eigenvalues. Prove that $(T - 5I)^{\dim V - 1}(T - 6I)^{\dim V - 1} = 0$.

5B.25 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and U is a subspace of V that is invariant under T.

(a) Prove that the minimal polynomial of T is a polynomial multiple of the minimal polynomial of the quotient operator T/U.

(b) Prove that

$$(\text{minimal polynomial of } T|_U) \times (\text{minimal polynomial of } T/U)$$

is a polynomial multiple of the minimal polynomial of T.

The quotient operator T/U was defined in Exercise 38 in Section 5A.

5B.26 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and U is a subspace of V that is invariant under T. Prove that the set of eigenvalues of T equals the union of the set of eigenvalues of $T|_U$ and the set of eigenvalues of T/U.

5B.27 Suppose $\mathbb{F} = \mathbb{R}$, V is finite-dimensional, and $T \in \text{End}(V)$. Prove that the minimal polynomial of $T_{\mathbb{C}}$ equals the minimal polynomial of T.

The complexification $T_{\mathbb{C}}$ was defined in Exercise 33 of Section 3B.

5B.28 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that the minimal polynomial of $T' \in \text{End}(V')$ equals the minimal polynomial of T.

The dual map T' was defined in Section 3F.

5B.29 Show that every operator on a finite-dimensional vector space of dimension at least two has an invariant subspace of dimension two.

Exercise 6 in Section 5C will give an improvement of this result when $\mathbb{F} = \mathbb{C}$.

5C Upper-Triangular Matrices

In Chapter 3 we defined the matrix of a linear map from a finite-dimensional vector space to another finite-dimensional vector space. That matrix depends on a choice of 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.

Definition 5.35

Matrix of an operator, $\text{Mat}(T)$

Suppose $T \in \text{End}(V)$. The *matrix* of T with respect to a basis $v_1, \dots, v_n$ of V is the n-by-n matrix

$$\text{Mat}(T) = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{pmatrix}$$

whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}v_1 + \cdots + A_{n,k}v_n.$$

Definition 5.35 *continued*

The notation $\text{Mat}\big(T, (v_1, \ldots, v_n)\big)$ is used if the basis is not clear from the context.

Operators have square matrices (meaning that the number of rows equals the number of columns), rather than the more general rectangular matrices that we considered earlier for linear maps.

If T is an operator on $\mathbb{F}^n$ and no basis is specified, assume that the basis in question is the standard one (where the kth basis vector is 1 in the kth slot and 0 in all other slots). You can then think of the kth column of $\text{Mat}(T)$ as T applied to the kth basis vector, where we identify n-by-1 column vectors with elements of $\mathbb{F}^n$.

The kth column of the matrix $\text{Mat}(T)$ is formed from the coefficients used to write Tv_k as a linear combination of the basis $v_1, \ldots, v_n$.

Example 5.36 **Matrix of an operator with respect to standard basis**

Define $T \in \text{End}(\mathbb{F}^3)$ by $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. Then the matrix of T with respect to the standard basis of $\mathbb{F}^3$ is

$$\text{Mat}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix},$$

as you should verify.

A central goal of linear algebra is to show that given an operator T on a finite-dimensional vector space 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 $\text{Mat}(T)$ 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 $*$ denotes the entries in all columns other than the first column. To prove this, let λ be an eigenvalue of T (one exists by 5.19) 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.

Diagonal of a matrix Definition 5.37

The *diagonal* of a square matrix consists of the entries on the line from the upper left corner to the bottom right corner.

For example, the diagonal of the matrix

$$\mathrm{Mat}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}$$

from Example 5.36 consists of the entries $2, 5, 8$, which are shown in red in the matrix above.

Upper-triangular matrix — Definition 5.38

A square matrix is called *upper triangular* if all entries below the diagonal are 0.

For example, the 3-by-3 matrix above 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—if n is large, then at least almost half the entries in an n-by-n upper-triangular matrix are 0.

We often use $*$ to denote matrix entries that we do not know or that are irrelevant to the questions being discussed.

The next result provides a useful connection between upper-triangular matrices and invariant subspaces.

Conditions for upper-triangular matrix — Result 5.39

Suppose $T \in \mathrm{End}(V)$ and $v_1, \dots, v_n$ is a basis of V. Then the following are equivalent.

(a) The matrix of T with respect to $v_1, \dots, v_n$ is upper triangular.

(b) $\mathrm{span}(v_1, \dots, v_k)$ is invariant under T for each $k = 1, \dots, n$.

(c) $Tv_k \in \mathrm{span}(v_1, \dots, v_k)$ for each $k = 1, \dots, n$.

Proof. First suppose (a) holds. To prove that (b) holds, suppose $k \in \{1, \dots, n\}$. If $j \in \{1, \dots, n\}$, then

$$Tv_j \in \mathrm{span}(v_1, \dots, v_j)$$

because the matrix of T with respect to $v_1, \dots, v_n$ is upper triangular. Because $\mathrm{span}(v_1, \dots, v_j) \subseteq \mathrm{span}(v_1, \dots, v_k)$ if $j \le k$, we see that

$$Tv_j \in \mathrm{span}(v_1, \dots, v_k)$$

for each $j \in \{1, \dots, k\}$. Thus $\mathrm{span}(v_1, \dots, v_k)$ is invariant under T, completing the proof that (a) implies (b).

Now suppose (b) holds, so $\mathrm{span}(v_1, \dots, v_k)$ is invariant under T for each $k = 1, \dots, n$. In particular, $Tv_k \in \mathrm{span}(v_1, \dots, v_k)$ for each $k = 1, \dots, n$. Thus (b) implies (c).

Now suppose (c) holds, so $Tv_k \in \operatorname{span}(v_1, \ldots, v_k)$ for each $k = 1, \ldots, n$. This means that when writing each Tv_k as a linear combination of the basis vectors $v_1, \ldots, v_n$, we need to use only the vectors $v_1, \ldots, v_k$. Hence all entries under the diagonal of $\operatorname{Mat}(T)$ are 0. Thus $\operatorname{Mat}(T)$ is an upper-triangular matrix, completing the proof that (c) implies (a).

We have shown that (a) $\implies$ (b) $\implies$ (c) $\implies$ (a), which shows that (a), (b), and (c) are equivalent. □

The next result tells us that if $T \in \operatorname{End}(V)$ and with respect to some basis of V we have

$$\operatorname{Mat}(T) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

then T satisfies a simple equation depending on $\lambda_1, \ldots, \lambda_n$.

Equation satisfied by operator with upper-triangular matrix — Result 5.40

Suppose $T \in \operatorname{End}(V)$ and V has a basis with respect to which T has an upper-triangular matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0.$$

Proof. Let $v_1, \ldots, v_n$ denote a basis of V with respect to which T has an upper-triangular matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then $Tv_1 = \lambda_1 v_1$, which means that $(T - \lambda_1 I)v_1 = 0$, which implies that $(T - \lambda_1 I) \cdots (T - \lambda_m I)v_1 = 0$ for $m = 1, \ldots, n$ (using the commutativity of each $T - \lambda_j I$ with each $T - \lambda_k I$).

Note that $(T - \lambda_2 I)v_2 \in \operatorname{span}(v_1)$. Thus $(T - \lambda_1 I)(T - \lambda_2 I)v_2 = 0$ (by the previous paragraph), which implies that $(T - \lambda_1 I) \cdots (T - \lambda_m I)v_2 = 0$ for $m = 2, \ldots, n$ (using the commutativity of each $T - \lambda_j I$ with each $T - \lambda_k I$).

Note that $(T - \lambda_3 I)v_3 \in \operatorname{span}(v_1, v_2)$. Thus by the previous paragraph, $(T - \lambda_1 I)(T - \lambda_2 I)(T - \lambda_3 I)v_3 = 0$, which implies that $(T - \lambda_1 I) \cdots (T - \lambda_m I)v_3 = 0$ for $m = 3, \ldots, n$ (using the commutativity of each $T - \lambda_j I$ with each $T - \lambda_k I$).

Continuing this pattern, we see that $(T - \lambda_1 I) \cdots (T - \lambda_n I)v_k = 0$ for each $k = 1, \ldots, n$. Thus $(T - \lambda_1 I) \cdots (T - \lambda_n I)$ is the 0 operator because it is 0 on each vector in a basis of V. □

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 next result shows.

Determination of eigenvalues from upper-triangular matrix — Result 5.41

Suppose $T \in \operatorname{End}(V)$ 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

$$\mathrm{Mat}(T) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Because $Tv_1 = \lambda_1 v_1$, we see that λ_1 is an eigenvalue of T.

Suppose $k \in \{2, \dots, n\}$. Then $(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$. Thus $T - \lambda_k I$ maps $\operatorname{span}(v_1, \dots, v_k)$ into $\operatorname{span}(v_1, \dots, v_{k-1})$. Because

$$\dim \operatorname{span}(v_1, \dots, v_k) = k \quad \text{and} \quad \dim \operatorname{span}(v_1, \dots, v_{k-1}) = k - 1,$$

this implies that $T - \lambda_k I$ restricted to $\operatorname{span}(v_1, \dots, v_k)$ is not injective (by 3.22). Thus there exists $v \in \operatorname{span}(v_1, \dots, v_k)$ such that $v \neq 0$ and $(T - \lambda_k I)v = 0$. Thus λ_k is an eigenvalue of T. Hence we have shown that every entry on the diagonal of $\mathrm{Mat}(T)$ is an eigenvalue of T.

To prove T has no other eigenvalues, let q be the polynomial defined by $q(z) = (z - \lambda_1) \cdots (z - \lambda_n)$. Then $q(T) = 0$ (by 5.40). Hence q is a polynomial multiple of the minimal polynomial of T (by 5.29). Thus every zero of the minimal polynomial of T is a zero of q. Because the zeros of the minimal polynomial of T are the eigenvalues of T (by 5.27), this implies that every eigenvalue of T is a zero of q. Hence the eigenvalues of T are all contained in the list $\lambda_1, \dots, \lambda_n$. □

Example 5.42 **Eigenvalues via an upper-triangular matrix**

Define $T \in \mathrm{End}(\mathbb{F}^3)$ by $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. The matrix of T with respect to the standard basis is

$$\mathrm{Mat}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

Now 5.41 implies that the eigenvalues of T are 2, 5, and 8.

The next example illustrates 5.44: an operator has an upper-triangular matrix with respect to some basis if and only if the minimal polynomial of the operator is the product of polynomials of degree 1.

Example 5.43 **Whether T has an upper-triangular matrix can depend on $\mathbb{F}$**

Define $T \in \mathrm{End}(\mathbb{F}^4)$ by

$$T(z_1, z_2, z_3, z_4) = (-z_2, z_1, 2z_1 + 3z_3, z_3 + 3z_4).$$

Example 5.43
continued

Thus with respect to the standard basis of $\mathbb{F}^4$, the matrix of T is

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

You can ask a computer to verify that the minimal polynomial of T is the polynomial p defined by

$$p(z) = 9 - 6z + 10z^2 - 6z^3 + z^4.$$

First consider the case $\mathbb{F} = \mathbb{R}$. Then the polynomial p factors as

$$p(z) = (z^2 + 1)(z - 3)(z - 3),$$

with no further factorization of $z^2 + 1$ as the product of two polynomials of degree 1 with real coefficients. Thus 5.44 states that there does not exist a basis of $\mathbb{R}^4$ with respect to which T has an upper-triangular matrix.

Now consider the case $\mathbb{F} = \mathbb{C}$. Then the polynomial p factors as

$$p(z) = (z - i)(z + i)(z - 3)(z - 3),$$

where all factors above have the form $z - \lambda_k$. Thus 5.44 states that there is a basis of $\mathbb{C}^4$ with respect to which T has an upper-triangular matrix. Indeed, you can verify that with respect to the basis $(4 - 3i, -3 - 4i, -3 + i, 1)$, $(4 + 3i, -3 + 4i, -3 - i, 1)$, $(0, 0, 0, 1)$, $(0, 0, 1, 0)$ of $\mathbb{C}^4$, the operator T has the upper-triangular matrix

$$\begin{pmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Result 5.44

Necessary and sufficient condition to have an upper-triangular matrix
Suppose V is finite-dimensional and $T \in \text{End}(V)$. Then T has an upper-triangular matrix with respect to some basis of V if and only if the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof. First suppose T has an upper-triangular matrix with respect to some basis of V. Let $\alpha_1, \ldots, \alpha_n$ denote the diagonal entries of that matrix. Define a polynomial $q \in \mathbb{F}[X]$ by

$$q(z) = (z - \alpha_1) \cdots (z - \alpha_n).$$

Then $q(T) = 0$, by 5.40. Hence q is a polynomial multiple of the minimal polynomial of T, by 5.29. Thus the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ with $\{\lambda_1, \ldots, \lambda_m\} \subseteq \{\alpha_1, \ldots, \alpha_n\}$.

To prove the implication in the other direction, now suppose the minimal polynomial of T equals $(z-\lambda_1)\cdots(z-\lambda_m)$ for some $\lambda_1,\ldots,\lambda_m \in \mathbb{F}$. We will use induction on m. To get started, if $m = 1$ then $z-\lambda_1$ is the minimal polynomial of T, which implies that $T = \lambda_1 I$, which implies that the matrix of T (with respect to any basis of V) is upper triangular.

Now suppose $m > 1$ and the desired result holds for all smaller positive integers. Let

$$U = \operatorname{im}(T-\lambda_m I).$$

Then U is invariant under T [this is a special case of 5.18 with $p(z) = z-\lambda_m$]. Thus $T|_U$ is an operator on U.

If $u \in U$, then $u = (T-\lambda_m I)v$ for some $v \in V$ and

$$(T-\lambda_1 I)\cdots(T-\lambda_{m-1}I)u = (T-\lambda_1 I)\cdots(T-\lambda_m I)v = 0.$$

Hence $(z-\lambda_1)\cdots(z-\lambda_{m-1})$ is a polynomial multiple of the minimal polynomial of $T|_U$, by 5.29. Thus the minimal polynomial of $T|_U$ is the product of at most $m-1$ terms of the form $z-\lambda_k$.

By our induction hypothesis, there is a basis $u_1,\ldots,u_M$ of U with respect to which $T|_U$ has an upper-triangular matrix. Thus for each $k \in \{1,\ldots,M\}$, we have (using 5.39)

$$Tu_k = (T|_U)(u_k) \in \operatorname{span}(u_1,\ldots,u_k). \tag{5.45}$$

Extend $u_1,\ldots,u_M$ to a basis $u_1,\ldots,u_M,v_1,\ldots,v_N$ of V. If $k \in \{1,\ldots,N\}$, then

$$Tv_k = (T-\lambda_m I)v_k + \lambda_m v_k.$$

The definition of U shows that $(T-\lambda_m I)v_k \in U = \operatorname{span}(u_1,\ldots,u_M)$. Thus the equation above shows that

$$Tv_k \in \operatorname{span}(u_1,\ldots,u_M,v_1,\ldots,v_k). \tag{5.46}$$

From 5.45 and 5.46, we conclude (using 5.39) that T has an upper-triangular matrix with respect to the basis $u_1,\ldots,u_M,v_1,\ldots,v_N$ of V, as desired. □

The set of numbers $\{\lambda_1,\ldots,\lambda_m\}$ from the previous result equals the set of eigenvalues of T (because the set of zeros of the minimal polynomial of T equals the set of eigenvalues of T, by 5.27), although the list $\lambda_1,\ldots,\lambda_m$ in the previous result may contain repetitions.

In Chapter 8 we will improve even the wonderful result below; see 8.37 and 8.46.

If $\mathbb{F} = \mathbb{C}$, then every operator on V has an upper-triangular matrix **Result 5.47**

Suppose V is a finite-dimensional complex vector space and $T \in \operatorname{End}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

Proof. The desired result follows immediately from 5.44 and the second version of the fundamental theorem of algebra (see 4.13). □

For an extension of the result above to two operators S and T such that

$$ST = TS,$$

see 5.80. Also, for an extension to more than two operators, see Exercise 9(b) in Section 5E.

Caution: If an operator $T \in \text{End}(V)$ has an upper-triangular matrix with respect to some basis $v_1, \dots, v_n$ of V, then the eigenvalues of T are exactly the entries on the diagonal of $\text{Mat}(T)$, as shown by 5.41, and furthermore v_1 is an eigenvector of T. However, $v_2, \dots, v_n$ need not be eigenvectors of T. Indeed, a basis vector v_k is an eigenvector of T if and only if all entries in the kth column of the matrix of T are 0, except possibly the kth entry.

You may recall from a previous course that every matrix of numbers can be changed to a matrix in what is called row echelon form. If one begins with a square matrix, the matrix in row echelon form will be an upper-triangular matrix. Do not confuse this upper-triangular matrix with the upper-triangular matrix of an operator with respect to some basis whose existence is proclaimed by 5.47 (if $\mathbb{F} = \mathbb{C}$)—there is no connection between these upper-triangular matrices.

The row echelon form of the matrix of an operator does not give us a list of the eigenvalues of the operator. In contrast, an upper-triangular matrix with respect to some basis gives us a list of all the eigenvalues of the operator. However, there is no method for computing exactly such an upper-triangular matrix, even though 5.47 guarantees its existence if $\mathbb{F} = \mathbb{C}$.

Exercises

5C.1 Prove or give a counterexample: If $T \in \text{End}(V)$ and T^2 has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some basis of V.

5C.2 Suppose A and B are upper-triangular matrices of the same size, with $\alpha_1, \dots, \alpha_n$ on the diagonal of A and $\beta_1, \dots, \beta_n$ on the diagonal of B.

(a) Show that $A + B$ is an upper-triangular matrix with $\alpha_1 + \beta_1, \dots, \alpha_n + \beta_n$ on the diagonal.

(b) Show that AB is an upper-triangular matrix with $\alpha_1\beta_1, \dots, \alpha_n\beta_n$ on the diagonal.

The results in this exercise are used in the proof of 5.81.

5C.3 Suppose $T \in \text{End}(V)$ is invertible and $v_1, \dots, v_n$ is a basis of V with respect to which the matrix of T is upper triangular, with $\lambda_1, \dots, \lambda_n$ on the diagonal. Show that the matrix of T^{-1} is also upper triangular with respect to the basis $v_1, \dots, v_n$, with

$$\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}$$

on the diagonal.

5C.4 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.

This exercise and the exercise below show that 5.41 fails without the hypothesis that an upper-triangular matrix is under consideration.

5C.5 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.

5C.6 Suppose $\mathbb{F} = \mathbb{C}$, V is finite-dimensional, and $T \in \text{End}(V)$. Prove that if $k \in \{1, \dots, \dim V\}$, then V has a k-dimensional subspace invariant under T.

5C.7 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and $v \in V$.

(a) Prove that there exists a unique monic polynomial p_v of smallest degree such that $p_v(T)v = 0$.

(b) Prove that the minimal polynomial of T is a polynomial multiple of p_v.

5C.8 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and there exists a nonzero vector $v \in V$ such that $T^2v + 2Tv = -2v$.

(a) Prove that if $\mathbb{F} = \mathbb{R}$, then there does not exist a basis of V with respect to which T has an upper-triangular matrix.

(b) Prove that if $\mathbb{F} = \mathbb{C}$ and A is an upper-triangular matrix that equals the matrix of T with respect to some basis of V, then $-1 + i$ or $-1 - i$ appears on the diagonal of A.

5C.9 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.

5C.10 Suppose $T \in \text{End}(V)$ and $v_1, \dots, v_n$ is a basis of V. Show that the following are equivalent.

(a) The matrix of T with respect to $v_1, \dots, v_n$ is lower triangular.

(b) $\text{span}(v_k, \dots, v_n)$ is invariant under T for each $k = 1, \dots, n$.

(c) $Tv_k \in \text{span}(v_k, \dots, v_n)$ for each $k = 1, \dots, n$.

A square matrix is called ***lower triangular*** *if all entries above the diagonal are* 0.

5C.11 Suppose $\mathbb{F} = \mathbb{C}$ and V is finite-dimensional. Prove that if $T \in \text{End}(V)$, then there exists a basis of V with respect to which T has a lower-triangular matrix.

5C.12 Suppose V is finite-dimensional, $T \in \text{End}(V)$ has an upper-triangular matrix with respect to some basis of V, and U is a subspace of V that is invariant under T.

(a) Prove that $T|_U$ has an upper-triangular matrix with respect to some basis of U.

(b) Prove that the quotient operator T/U has an upper-triangular matrix with respect to some basis of V/U.

The quotient operator T/U was defined in Exercise 38 in Section 5A.

5C.13 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Suppose there exists a subspace U of V that is invariant under T such that $T|_U$ has an upper-triangular matrix with respect to some basis of U and also T/U has an upper-triangular matrix with respect to some basis of V/U. Prove that T has an upper-triangular matrix with respect to some basis of V.

5C.14 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that T has an upper-triangular matrix with respect to some basis of V if and only if the dual operator T' has an upper-triangular matrix with respect to some basis of the dual space V'.

5D Diagonalizable Operators

Diagonal Matrices

Diagonal matrix
A *diagonal matrix* is a square matrix that is 0 everywhere except possibly on the diagonal.

Definition 5.48

Example 5.49 **Diagonal matrix**

$$\begin{pmatrix} 8 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

is a diagonal matrix.

Every diagonal matrix is upper triangular. Diagonal matrices typically have many more 0's than most upper-triangular matrices of the same size.

If an operator has a diagonal matrix with respect to some basis, then the entries on the diagonal are precisely the eigenvalues of the operator; this follows from 5.41 (or find an easier direct proof for diagonal matrices).

Diagonalizable
An operator on V is called *diagonalizable* if the operator has a diagonal matrix with respect to some basis of V.

Definition 5.50

Example 5.51 **Diagonalization may require a different basis**

Define $T \in \text{End}(\mathbb{R}^2)$ by

$$T(x, y) = (41x + 7y, -20x + 74y).$$

Example 5.51 *continued*

The matrix of T with respect to the standard basis of $\mathbb{R}^2$ is

$$\begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix},$$

which is not a diagonal matrix. However, T is diagonalizable. Specifically, the matrix of T with respect to the basis $(1, 4), (7, 5)$ is

$$\begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix}$$

because $T(1, 4) = (69, 276) = 69(1, 4)$ and $T(7, 5) = (322, 230) = 46(7, 5)$.

For $\lambda \in \mathbb{F}$, we will find it convenient to have a name and a notation for the set of vectors that an operator T maps to λ times the vector.

Eigenspace, $E(\lambda, T)$ Definition 5.52

Suppose $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$. The *eigenspace* of T corresponding to λ is the subspace $E(\lambda, T)$ of V defined by

$$E(\lambda, T) = \ker(T - \lambda I) = \{v \in V : Tv = \lambda v\}.$$

Hence $E(\lambda, T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

For $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$, the set $E(\lambda, T)$ 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 $E(\lambda, T) \neq \{0\}$.

Example 5.53 **Eigenspaces of an operator**

Suppose the matrix of an operator $T \in \text{End}(V)$ with respect to a basis v_1, v_2, v_3 of V is the matrix in Example 5.49. Then

$$E(8, T) = \text{span}(v_1), \qquad E(5, T) = \text{span}(v_2, v_3).$$

If λ is an eigenvalue of an operator $T \in \text{End}(V)$, then T restricted to $E(\lambda, T)$ is just the operator of multiplication by λ.

Sum of eigenspaces is a direct sum Result 5.54

Suppose $T \in \text{End}(V)$ and $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Furthermore, if V is finite-dimensional, then

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

Proof. To show that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, suppose

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

where each v_k is in $E(\lambda_k, T)$. Because eigenvectors corresponding to distinct eigenvalues are linearly independent (by 5.11), this implies that each v_k equals 0. Thus $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum (by 1.45), as desired.

Now suppose V is finite-dimensional. Then

$$\begin{aligned}\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) &= \dim\big(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)\big)\\ &\le \dim V,\end{aligned}$$

where the first line follows from 3.94 and the second line follows from 2.37. □

Conditions for Diagonalizability

The following characterizations of diagonalizable operators will be useful.

Result 5.55

Conditions equivalent to diagonalizability

Suppose V is finite-dimensional and $T \in \text{End}(V)$. Let $\lambda_1, \ldots, \lambda_m$ 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) $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.

(d) $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Proof. An operator $T \in \text{End}(V)$ has a diagonal matrix

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

with respect to a basis $v_1, \ldots, v_n$ of V if and only if $Tv_k = \lambda_k v_k$ for each k. Thus (a) and (b) are equivalent.

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 5.54 shows that (c) holds, proving that (b) implies (c).

That (c) implies (d) follows immediately from 3.94.

Finally, suppose (d) holds; thus

$$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T). \tag{5.56}$$

Choose a basis of each $E(\lambda_k, T)$; put all these bases together to form a list $v_1, \dots, v_n$ of eigenvectors of T, where $n = \dim V$ (by 5.56). To show that this list is linearly independent, suppose

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

where $a_1, \dots, a_n \in \mathbb{F}$. For each $k = 1, \dots, m$, let u_k denote the sum of all the terms $a_j v_j$ such that $v_j \in E(\lambda_k, T)$. Thus each u_k is in $E(\lambda_k, T)$, and

$$u_1 + \cdots + u_m = 0.$$

Because eigenvectors corresponding to distinct eigenvalues are linearly independent (see 5.11), this implies that each u_k equals 0. Because each u_k is a sum of terms $a_j v_j$, where the v_j's were chosen to be a basis of $E(\lambda_k, T)$, this implies that all a_j's equal 0. Thus $v_1, \dots, v_n$ is linearly independent and hence is a basis of V (by 2.38). Thus (d) implies (b), completing the proof. □

For additional conditions equivalent to diagonalizability, see 5.62, Exercises 5 and 15 in this section, Exercise 24 in Section 7B, and Exercise 15 in Section 8A.

As we know, every operator on a nonzero finite-dimensional complex vector space has an eigenvalue. However, not every operator on a nonzero finite-dimensional complex vector space has enough eigenvectors to be diagonalizable, as shown by the next example.

Example 5.57 **An operator that is not diagonalizable**

Define an operator $T \in \text{End}(\mathbb{F}^3)$ by $T(a, b, c) = (b, c, 0)$. The matrix of T with respect to the standard basis of $\mathbb{F}^3$ is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

which is an upper-triangular matrix but is not a diagonal matrix.

As you should verify, 0 is the only eigenvalue of T and furthermore

$$E(0, T) = \{(a, 0, 0) \in \mathbb{F}^3 : a \in \mathbb{F}\}.$$

Hence conditions (b), (c), and (d) of 5.55 fail (of course, because these conditions are equivalent, it is sufficient to check that only one of them fails). Thus condition (a) of 5.55 also fails. Hence T is not diagonalizable, regardless of whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$.

The next result shows that if an operator has as many distinct eigenvalues as the dimension of its domain, then the operator is diagonalizable.

Result 5.58 **Enough eigenvalues implies diagonalizability**

Suppose V is finite-dimensional and $T \in \text{End}(V)$ has $\dim V$ distinct eigenvalues. Then T is diagonalizable.

Proof. Suppose T has distinct eigenvalues $\lambda_1, \dots, \lambda_{\dim V}$. For each k, let $v_k \in V$ be an eigenvector corresponding to the eigenvalue λ_k. Because eigenvectors corresponding to distinct eigenvalues are linearly independent (see 5.11), $v_1, \dots, v_{\dim V}$ is linearly independent.

A linearly independent list of $\dim V$ vectors in V is a basis of V (see 2.38); 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. □

In later chapters we will find additional conditions that imply that certain operators are diagonalizable. For example, see the real spectral theorem (7.29) and the complex spectral theorem (7.31).

The result above gives a sufficient condition for an operator to be diagonalizable. However, this condition is not necessary. For example, the operator T on $\mathbb{F}^3$ defined by $T(x, y, z) = (6x, 6y, 7z)$ has only two eigenvalues (6 and 7) and $\dim \mathbb{F}^3 = 3$, but T is diagonalizable (by the standard basis of $\mathbb{F}^3$).

The next example illustrates the importance of diagonalization, which can be used to compute high powers of an operator, taking advantage of the equation $T^k v = \lambda^k v$ if v is an eigenvector of T with eigenvalue λ.

For a spectacular application of these techniques, see Exercise 21, which shows how to use diagonalization to find an exact formula for the nth term of the Fibonacci sequence.

Example 5.59 **Using diagonalization to compute T^{100}**

Define $T \in \text{End}(\mathbb{F}^3)$ by $T(x, y, z) = (2x + y, 5y + 3z, 8z)$. 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 an upper-triangular matrix but it is not a diagonal matrix. By 5.41, the eigenvalues of T are 2, 5, and 8. Because T is an operator on a vector space of dimension three and T has three distinct eigenvalues, 5.58 assures us that there exists a basis of $\mathbb{F}^3$ 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 $\lambda = 8$. Solving these simple equations shows that for $\lambda = 2$ we have an eigenvector $(1, 0, 0)$, for $\lambda = 5$ we have an eigenvector $(1, 3, 0)$, and for $\lambda = 8$ we have an eigenvector $(1, 6, 6)$.

Thus $(1, 0, 0)$, $(1, 3, 0)$, $(1, 6, 6)$ is a basis of $\mathbb{F}^3$ consisting of eigenvectors of T, and with respect to this basis the matrix of T is the diagonal matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 8 \end{pmatrix}.$$

Example 5.59 *continued*

To compute $T^{100}(0,0,1)$, for example, write $(0,0,1)$ as a linear combination of our basis of eigenvectors:

$$(0,0,1) = (1/6)(1,0,0) - (1/3)(1,3,0) + (1/6)(1,6,6).$$

Now apply T^{100} to both sides of the equation above, getting

$$\begin{aligned} T^{100}(0,0,1) &= 1/6\big(T^{100}(1,0,0)\big) - 1/3\big(T^{100}(1,3,0)\big) + 1/6\big(T^{100}(1,6,6)\big) \\ &= 1/6\big(2^{100}(1,0,0) - 2\cdot 5^{100}(1,3,0) + 8^{100}(1,6,6)\big) \\ &= 1/6\big(2^{100} - 2\cdot 5^{100} + 8^{100},\ 6\cdot 8^{100} - 6\cdot 5^{100},\ 6\cdot 8^{100}\big). \end{aligned}$$

We saw earlier that an operator T on a finite-dimensional vector space V has an upper-triangular matrix with respect to some basis of V if and only if the minimal polynomial of T equals $(z-\lambda_1)\cdots(z-\lambda_m)$ for some $\lambda_1,\ldots,\lambda_m \in \mathbb{F}$ (see 5.44). As we previously noted (see 5.47), this condition is always satisfied if $\mathbb{F} = \mathbb{C}$.

Our next result 5.62 states that an operator $T \in \text{End}(V)$ has a diagonal matrix with respect to some basis of V if and only if the minimal polynomial of T equals $(z-\lambda_1)\cdots(z-\lambda_m)$ for some ***distinct*** $\lambda_1,\ldots,\lambda_m \in \mathbb{F}$. Before formally stating this result, we give two examples of using it.

Example 5.60 **Diagonalizable, but with no known exact eigenvalues**

Define $T \in \text{End}(\mathbb{C}^5)$ by

$$T(z_1,z_2,z_3,z_4,z_5) = (-3z_5, z_1 + 6z_5, z_2, z_3, z_4).$$

The matrix of T is shown in Example 5.26, where we showed that the minimal polynomial of T is $3 - 6z + z^5$.

As mentioned in Example 5.28, no exact expression is known for any of the zeros of this polynomial, but numerical techniques show that the zeros of this polynomial are approximately -1.67, 0.51, 1.40, $-0.12 + 1.59i$, $-0.12 - 1.59i$.

The software that produces these approximations is accurate to more than three digits. Thus these approximations are good enough to show that the five numbers above are distinct. The minimal polynomial of T equals the fifth degree monic polynomial with these zeros. Now 5.62 shows that T is diagonalizable.

Example 5.61 **Showing that an operator is not diagonalizable**

Define $T \in \text{End}(\mathbb{F}^3)$ by

$$T(z_1,z_2,z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3).$$

Example 5.61 *continued*

The matrix of T with respect to the standard basis of $\mathbb{F}^3$ is

$$\begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

The matrix above is an upper-triangular matrix but is not a diagonal matrix. Might T have a diagonal matrix with respect to some other basis of $\mathbb{F}^3$?

To answer this question, we will find the minimal polynomial of T. First note that the eigenvalues of T are the diagonal entries of the matrix above (by 5.41). Thus the zeros of the minimal polynomial of T are $6, 7$ [by 5.27(a)]. The diagonal of the matrix above tells us that $(T - 6I)^2(T - 7I) = 0$ (by 5.40). The minimal polynomial of T has degree at most 3 (by 5.22). Putting all this together, we see that 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 $(z - 6)^2(z - 7)$.

Now 5.62 shows that T is not diagonalizable.

Result 5.62

Necessary and sufficient condition for diagonalizability
Suppose V is finite-dimensional and $T \in \text{End}(V)$. Then T is diagonalizable if and only if the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some list of distinct numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof. First suppose T is diagonalizable. Thus there is a basis $v_1, \ldots, v_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then for each v_j, there exists λ_k with $(T - \lambda_k I)v_j = 0$. Thus

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)v_j = 0,$$

which implies that the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$.

To prove the implication in the other direction, now suppose the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some list of distinct numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. Thus

$$(T - \lambda_1 I) \cdots (T - \lambda_m I) = 0. \tag{5.63}$$

We will prove that T is diagonalizable by induction on m. To get started, suppose $m = 1$. Then $T - \lambda_1 I = 0$, which means that T is a scalar multiple of the identity operator, which implies that T is diagonalizable.

Now suppose that $m > 1$ and the desired result holds for all smaller values of m. The subspace $\text{im}(T - \lambda_m I)$ is invariant under T [this is a special case of 5.18 with $p(z) = z - \lambda_m$]. Thus T restricted to $\text{im}(T - \lambda_m I)$ is an operator on $\text{im}(T - \lambda_m I)$.

If $u \in \text{im}(T - \lambda_m I)$, then $u = (T - \lambda_m I)v$ for some $v \in V$, and 5.63 implies

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_m I)v = 0. \tag{5.64}$$

Hence $(z-\lambda_1)\cdots(z-\lambda_{m-1})$ is a polynomial multiple of the minimal polynomial of T restricted to $\operatorname{im}(T-\lambda_m I)$ [by 5.29]. Thus by our induction hypothesis, there is a basis of $\operatorname{im}(T-\lambda_m I)$ consisting of eigenvectors of T.

Suppose that $u \in \operatorname{im}(T-\lambda_m I) \cap \ker(T-\lambda_m I)$. Then $Tu = \lambda_m u$. Now 5.64 implies that

$$\begin{aligned} 0 &= (T-\lambda_1 I)\cdots(T-\lambda_{m-1}I)u \\ &= (\lambda_m-\lambda_1)\cdots(\lambda_m-\lambda_{m-1})u. \end{aligned}$$

Because $\lambda_1, \ldots, \lambda_m$ are distinct, the equation above implies that $u = 0$. Hence $\operatorname{im}(T-\lambda_m I) \cap \ker(T-\lambda_m I) = \{0\}$.

Thus $\operatorname{im}(T-\lambda_m I) + \ker(T-\lambda_m I)$ is a direct sum (by 1.46) whose dimension is $\dim V$ (by 3.94 and 3.21). Hence $\operatorname{im}(T-\lambda_m I) \oplus \ker(T-\lambda_m I) = V$. Every nonzero vector in $\ker(T-\lambda_m I)$ is an eigenvector of T with eigenvalue λ_m. Earlier in this proof we saw that there is a basis of $\operatorname{im}(T-\lambda_m I)$ consisting of eigenvectors of T. Adjoining to that basis a basis of $\ker(T-\lambda_m I)$ gives a basis of V consisting of eigenvectors of T. The matrix of T with respect to this basis is a diagonal matrix, as desired. □

No formula exists for the zeros of polynomials of degree 5 or greater. However, the previous result can be used to determine whether an operator on a complex vector space is diagonalizable without even finding approximations of the zeros of the minimal polynomial—see Exercise 15.

The next result will be a key tool when we prove a result about the simultaneous diagonalization of two operators; see 5.76. Note how the use of a characterization of diagonalizable operators in terms of the minimal polynomial (see 5.62) leads to a short proof of the next result.

Restriction of diagonalizable operator to invariant subspace — Result 5.65

Suppose $T \in \operatorname{End}(V)$ is diagonalizable and U is a subspace of V that is invariant under T. Then $T|_U$ is a diagonalizable operator on U.

Proof. Because the operator T is diagonalizable, the minimal polynomial of T equals $(z-\lambda_1)\cdots(z-\lambda_m)$ for some list of distinct numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ (by 5.62). The minimal polynomial of T is a polynomial multiple of the minimal polynomial of $T|_U$ (by 5.31). Hence the minimal polynomial of $T|_U$ has the form required by 5.62, which shows that $T|_U$ is diagonalizable. □

Gershgorin Disk Theorem

Gershgorin disks — Definition 5.66

Suppose $T \in \operatorname{End}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Let A denote the matrix of T with respect to this basis. A *Gershgorin disk* of T with respect to the basis

Definition 5.66 *continued*

$v_1, \dots, v_n$ is a set of the form

$$\Big\{z \in \mathbb{F} : |z - A_{j,j}| \le \sum_{\substack{k=1 \\ k \ne j}}^{n} |A_{j,k}|\Big\},$$

where $j \in \{1, \dots, n\}$.

Because there are n choices for j in the definition above, T has n Gershgorin disks. If $\mathbb{F} = \mathbb{C}$, then for each $j \in \{1, \dots, n\}$, the corresponding Gershgorin disk is a closed disk in $\mathbb{C}$ centered at $A_{j,j}$, which is the jth entry on the diagonal of A. The radius of this closed disk is the sum of the absolute values of the entries in row j of A, excluding the diagonal entry. If $\mathbb{F} = \mathbb{R}$, then the Gershgorin disks are closed intervals in $\mathbb{R}$.

In the special case that the square matrix A above is a diagonal matrix, each Gershgorin disk consists of a single point that is a diagonal entry of A (and each eigenvalue of T is one of those points, as required by the next result). One consequence of our next result is that if the nondiagonal entries of A are small, then each eigenvalue of T is near a diagonal entry of A.

Gershgorin disk theorem Result 5.67

Suppose $T \in \text{End}(V)$ and $v_1, \dots, v_n$ is a basis of V. Then each eigenvalue of T is contained in some Gershgorin disk of T with respect to the basis $v_1, \dots, v_n$.

Proof. Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of T. Let $w \in V$ be a corresponding eigenvector. There exist $c_1, \dots, c_n \in \mathbb{F}$ such that

$$w = c_1 v_1 + \cdots + c_n v_n. \tag{5.68}$$

Let A denote the matrix of T with respect to the basis $v_1, \dots, v_n$. Applying T to both sides of the equation above gives

$$\begin{aligned} \lambda w &= \sum_{k=1}^{n} c_k T v_k && (5.69) \\ &= \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j \\ &= \sum_{j=1}^{n} \Big(\sum_{k=1}^{n} A_{j,k} c_k \Big) v_j. && (5.70) \end{aligned}$$

Let $j \in \{1, \dots, n\}$ be such that

$$|c_j| = \max\{|c_1|, \dots, |c_n|\}.$$

Using 5.68, we see that the coefficient of v_j on the left side of 5.69 equals λc_j, which must equal the coefficient of v_j on the right side of 5.70. In other words,

$$\lambda c_j = \sum_{k=1}^{n} A_{j,k} c_k.$$

Subtract $A_{j,j}c_j$ from each side of the equation above and then divide both sides by c_j to get

$$\begin{aligned}|\lambda - A_{j,j}| &= \left|\sum_{\substack{k=1\\k\neq j}}^{n} A_{j,k}\frac{c_k}{c_j}\right| \\ &\leq \sum_{\substack{k=1\\k\neq j}}^{n} |A_{j,k}|.\end{aligned}$$

Thus λ is in the jth Gershgorin disk with respect to the basis $v_1, \ldots, v_n$. □

Exercise 22 gives a nice application of the Gershgorin disk theorem.

Exercise 23 states that the radius of each Gershgorin disk could be changed to the sum of the absolute values of corresponding column entries (instead of row entries), excluding the diagonal entry, and the theorem above would still hold.

The Gershgorin disk theorem is named for Semyon Aronovich Gershgorin, who published this result in 1931.

Exercises

5D.1 Suppose V is a finite-dimensional complex vector space and $T \in \text{End}(V)$.

(a) Prove that if $T^4 = I$, then T is diagonalizable.

(b) Prove that if $T^4 = T$, then T is diagonalizable.

(c) Give an example of an operator $T \in \text{End}(\mathbb{C}^2)$ such that $T^4 = T^2$ and T is not diagonalizable.

5D.2 Suppose $T \in \text{End}(V)$ has a diagonal matrix A with respect to some basis of V. Prove that if $\lambda \in \mathbb{F}$, then λ appears on the diagonal of A precisely $\dim E(\lambda, T)$ times.

5D.3 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that if the operator T is diagonalizable, then $V = \ker T \oplus \text{im}\, T$.

5D.4 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that the following are equivalent.

(a) $V = \ker T \oplus \text{im}\, T$.

(b) $V = \ker T + \text{im}\, T$.

(c) $\ker T \cap \text{im}\, T = \{0\}$.

5D.5 Suppose V is a finite-dimensional complex vector space and $T \in \text{End}(V)$. Prove that T is diagonalizable if and only if

$$V = \ker(T - \lambda I) \oplus \text{im}(T - \lambda I)$$

for every $\lambda \in \mathbb{C}$.

5D.6 Suppose $T \in \text{End}(\mathbb{F}^5)$ and $\dim E(8, T) = 4$. Prove that $T - 2I$ or $T - 6I$ is invertible.

5D.7 Suppose $T \in \text{End}(V)$ is invertible. Prove that

$$E(\lambda, T) = E\left(\tfrac{1}{\lambda}, T^{-1}\right)$$

for every $\lambda \in \mathbb{F}$ with $\lambda \neq 0$.

5D.8 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct nonzero eigenvalues of T. Prove that

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim \text{im}\, T.$$

5D.9 Suppose $R, T \in \text{End}(\mathbb{F}^3)$ each have $2, 6, 7$ as eigenvalues. Prove that there exists an invertible operator $S \in \text{End}(\mathbb{F}^3)$ such that $R = S^{-1}TS$.

5D.10 Find $R, T \in \text{End}(\mathbb{F}^4)$ 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 $S \in \text{End}(\mathbb{F}^4)$ such that $R = S^{-1}TS$.

5D.11 Find $T \in \text{End}(\mathbb{C}^3)$ 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 $\mathbb{C}^3$.

5D.12 Suppose $T \in \text{End}(\mathbb{C}^3)$ 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 $\mathbb{C}^3$. Prove that there exists $(z_1, z_2, z_3) \in \mathbb{C}^3$ such that

$$T(z_1, z_2, z_3) = (6 + 8z_1, 7 + 8z_2, 13 + 8z_3).$$

5D.13 Suppose A is a diagonal matrix with distinct entries on the diagonal and B is a matrix of the same size as A. Show that $AB = BA$ if and only if B is a diagonal matrix.

5D.14

(a) Give an example of a finite-dimensional complex vector space and an operator T on that vector space such that T^2 is diagonalizable but T is not diagonalizable.

(b) Suppose $\mathbb{F} = \mathbb{C}$, k is a positive integer, and $T \in \text{End}(V)$ is invertible. Prove that T is diagonalizable if and only if T^k is diagonalizable.

5D.15 Suppose V is a finite-dimensional complex vector space, $T \in \text{End}(V)$, and p is the minimal polynomial of T. Prove that the following are equivalent.

(a) T is diagonalizable.

(b) There does not exist $\lambda \in \mathbb{C}$ such that p is a polynomial multiple of $(z - \lambda)^2$.

(c) p and its derivative p' have no zeros in common.

(d) The greatest common divisor of p and p' is the constant polynomial 1.

The ***greatest common divisor*** *of p and p' is the monic polynomial q of largest degree such that p and p' are both polynomial multiples of q. The Euclidean algorithm for polynomials (look it up) can quickly determine the greatest common divisor of two polynomials, without requiring any information about the zeros of the polynomials. Thus the equivalence of (a) and (d) above shows that we can determine whether T is diagonalizable without knowing anything about the zeros of p.*

5D.16 Suppose that $T \in \text{End}(V)$ is diagonalizable. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Prove that a subspace U of V is invariant under T if and only if there exist subspaces $U_1, \ldots, U_m$ of V such that $U_k \subseteq E(\lambda_k, T)$ for each k and $U = U_1 \oplus \cdots \oplus U_m$.

5D.17 Suppose V is finite-dimensional. Prove that $\text{End}(V)$ has a basis consisting of diagonalizable operators.

5D.18 Suppose that $T \in \text{End}(V)$ is diagonalizable and U is a subspace of V that is invariant under T. Prove that the quotient operator T/U is a diagonalizable operator on V/U.

The quotient operator T/U was defined in Exercise 38 in Section 5A.

5D.19 Prove or give a counterexample: If $T \in \text{End}(V)$ and there exists a subspace U of V that is invariant under T such that $T|_U$ and T/U are both diagonalizable, then T is diagonalizable.

See Exercise 13 in Section 5C for an analogous statement about upper-triangular matrices.

5D.20 Suppose V is finite-dimensional and $T \in \text{End}(V)$. Prove that T is diagonalizable if and only if the dual operator T' is diagonalizable.

5D.21 The *Fibonacci sequence* $F_0, F_1, F_2, \ldots$ is defined by

$$F_0 = 0, \ F_1 = 1, \text{ and } F_n = F_{n-2} + F_{n-1} \text{ for } n \geq 2.$$

Define $T \in \text{End}(\mathbb{R}^2)$ by $T(x, y) = (y, x + y)$.

(a) Show that $T^n(0, 1) = (F_n, F_{n+1})$ for each nonnegative integer n.

(b) Find the eigenvalues of T.

(c) Find a basis of $\mathbb{R}^2$ consisting of eigenvectors of T.

(d) Use the solution to (c) to compute $T^n(0,1)$. Conclude that

$$F_n = 1/\sqrt{5}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

for each nonnegative integer n.

(e) Use (d) to conclude that if n is a nonnegative integer, then the Fibonacci number F_n is the integer that is closest to

$$1/\sqrt{5}\left(\frac{1+\sqrt{5}}{2}\right)^n.$$

Each F_n is a nonnegative integer, even though the right side of the formula in (d) does not look like an integer. The number

$$\frac{1+\sqrt{5}}{2}$$

is called the ***golden ratio****.*

5D.22 Suppose $T \in \text{End}(V)$ and A is an n-by-n matrix that is the matrix of T with respect to some basis of V. Prove that if

$$|A_{j,j}| > \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|$$

for each $j \in \{1, \ldots, n\}$, then T is invertible.

This exercise states that if the diagonal entries of the matrix of T are large compared to the nondiagonal entries, then T is invertible.

5D.23 Suppose the definition of the Gershgorin disks is changed so that the radius of the kth disk is the sum of the absolute values of the entries in column (instead of row) k of A, excluding the diagonal entry. Show that the Gershgorin disk theorem (5.67) still holds with this changed definition.

5E Commuting Operators

Commute — **Definition 5.71**

- Two operators S and T on the same vector space *commute* if $ST = TS$.
- Two square matrices A and B of the same size *commute* if $AB = BA$.

For example, if I is the identity operator on V and $\lambda \in \mathbb{F}$, then λI commutes with every operator on V.

As another example, if T is an operator then T^2 and T^3 commute. More generally, if $p, q \in \mathbb{F}[X]$, then $p(T)$ and $q(T)$ commute [see 5.17(b)].

Example 5.72 **Partial differentiation operators commute**

Suppose m is a nonnegative integer. Let $\mathbb{C}[w,z]_{\le m}$ denote the complex vector space of polynomials (with coefficients in $\mathbb{C}$) in two variables and of degree at most m, with the usual operations of addition and scalar multiplication of $\mathbb{C}$-valued functions. Thus the elements of $\mathbb{C}[w,z]_{\le m}$ are functions p from $\mathbb{C}^2$ to $\mathbb{C}$ of the form

$$p(w,z) = \sum_{j+k\le m} a_{j,k} w^j z^k, \tag{5.73}$$

where the indices j and k take on all nonnegative integer values such that $j+k \le m$, each $a_{j,k}$ is in $\mathbb{C}$, and $w^j z^k$ denotes the function on $\mathbb{C}^2$ defined by $(w,z) \mapsto w^j z^k$.

Define operators $D_w, D_z \in \operatorname{End}\big(\mathbb{C}[w,z]_{\le m}\big)$ by

$$D_w p = \frac{\partial p}{\partial w} = \sum_{j+k\le m} j a_{j,k} w^{j-1} z^k \quad \text{and} \quad D_z p = \frac{\partial p}{\partial z} = \sum_{j+k\le m} k a_{j,k} w^j z^{k-1},$$

where p is as in 5.73. The operators D_w and D_z are called partial differentiation operators because each of these operators differentiates with respect to one of the variables while pretending that the other variable is a constant.

The operators D_w and D_z commute because if p is as in 5.73, then

$$(D_w D_z)p = \sum_{j+k\le m} jk a_{j,k} w^{j-1} z^{k-1} = (D_z D_w)p.$$

The equation $D_w D_z = D_z D_w$ on $\mathbb{C}[w,z]_{\le m}$ illustrates a more general result that the order of partial differentiation does not matter for nice functions.

Commuting matrices are unusual. For example, there are 214,358,881 ordered pairs of 2-by-2 matrices all of whose entries are integers in the interval $[-5,5]$. Only about 0.3% of these ordered pairs of matrices commute.

All 214,358,881 (which equals 11^8) ordered pairs of the 2-by-2 matrices under consideration were checked by a computer to discover that only 674,609 of these ordered pairs of matrices commute.

The next result shows that two operators commute if and only if their matrices (with respect to the same basis) commute.

Commuting operators correspond to commuting matrices **Result 5.74**

Suppose $S, T \in \operatorname{End}(V)$ and $v_1, \dots, v_n$ is a basis of V. Then S and T commute if and only if $\operatorname{Mat}\big(S, (v_1, \dots, v_n)\big)$ and $\operatorname{Mat}\big(T, (v_1, \dots, v_n)\big)$ commute.

Proof. We have

$$\begin{aligned} S \text{ and } T \text{ commute} &\iff ST = TS \\ &\iff \operatorname{Mat}(ST) = \operatorname{Mat}(TS) \\ &\iff \operatorname{Mat}(S)\operatorname{Mat}(T) = \operatorname{Mat}(T)\operatorname{Mat}(S) \\ &\iff \operatorname{Mat}(S) \text{ and } \operatorname{Mat}(T) \text{ commute}, \end{aligned}$$

as desired. □

The next result shows that if two operators commute, then every eigenspace for one operator is invariant under the other operator. This result, which we will use several times, is one of the main reasons why a pair of commuting operators behaves better than a pair of operators that does not commute.

Eigenspace is invariant under commuting operator — Result 5.75

Suppose $S, T \in \text{End}(V)$ commute and $\lambda \in \mathbb{F}$. Then $E(\lambda, S)$ is invariant under T.

Proof. Suppose $v \in E(\lambda, S)$. Then

$$S(Tv) = (ST)v = (TS)v = T(Sv) = T(\lambda v) = \lambda Tv.$$

The equation above shows that $Tv \in E(\lambda, S)$. Thus $E(\lambda, S)$ is invariant under T. □

Suppose we have two operators, each of which is diagonalizable. If we want to do computations involving both operators (for example, involving their sum), then we want the two operators to be diagonalizable by the same basis, which according to the next result is possible when the two operators commute.

Simultaneous diagonalizability $\iff$ commutativity — Result 5.76

Two diagonalizable operators on the same vector space have diagonal matrices with respect to the same basis if and only if the two operators commute.

Proof. First suppose $S, T \in \text{End}(V)$ have diagonal matrices with respect to the same basis. The product of two diagonal matrices of the same size is the diagonal matrix obtained by multiplying the corresponding elements of the two diagonals. Thus any two diagonal matrices of the same size commute. Thus S and T commute, by 5.74.

To prove the implication in the other direction, now suppose that $S, T \in \text{End}(V)$ are diagonalizable operators that commute. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of S. Because S is diagonalizable, 5.55(c) shows that

$$V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S). \tag{5.77}$$

For each $k = 1, \dots, m$, the subspace $E(\lambda_k, S)$ is invariant under T (by 5.75). Because T is diagonalizable, 5.65 implies that $T|_{E(\lambda_k, S)}$ is diagonalizable for each k. Hence for each $k = 1, \dots, m$, there is a basis of $E(\lambda_k, S)$ consisting of eigenvectors of T. Putting these bases together gives a basis of V (because of 5.77), with each vector in this basis being an eigenvector of both S and T. Thus S and T both have diagonal matrices with respect to this basis, as desired. □

See Exercise 2 for an extension of the result above to more than two operators.

Suppose V is a finite-dimensional nonzero complex vector space. Then every operator on V has an eigenvector (see 5.19). The next result shows that if two operators on V commute, then there is a vector in V that is an eigenvector for both operators (but the two commuting operators might not have a common eigenvalue). For an extension of the next result to more than two operators, see Exercise 9(a).

Common eigenvector for commuting operators Result 5.78

Every pair of commuting operators on a finite-dimensional nonzero complex vector space has a common eigenvector.

Proof. Suppose V is a finite-dimensional nonzero complex vector space and $S, T \in \text{End}(V)$ commute. Let λ be an eigenvalue of S (5.19 tells us that S does indeed have an eigenvalue). Thus $E(\lambda, S) \neq \{0\}$. Also, $E(\lambda, S)$ is invariant under T (by 5.75).

Thus $T|_{E(\lambda,S)}$ has an eigenvector (again using 5.19), which is an eigenvector for both S and T, completing the proof. □

Example 5.79 **Common eigenvector for partial differentiation operators**

Let $\mathbb{C}[w,z]_{\leq m}$ be as in Example 5.72 and let $D_w, D_z \in \text{End}\big(\mathbb{C}[w,z]_{\leq m}\big)$ be the commuting partial differentiation operators in that example. As you can verify, 0 is the only eigenvalue of each of these operators. Also

$$E(0, D_w) = \left\{\sum_{k=0}^{m} a_k z^k : a_0, \ldots, a_m \in \mathbb{C}\right\},$$

$$E(0, D_z) = \left\{\sum_{j=0}^{m} c_j w^j : c_0, \ldots, c_m \in \mathbb{C}\right\}.$$

The intersection of these two eigenspaces is the set of common eigenvectors of the two operators. Because $E(0, D_w) \cap E(0, D_z)$ is the set of constant functions, we see that D_w and D_z indeed have a common eigenvector, as promised by 5.78.

The next result extends 5.47 (the existence of a basis that gives an upper-triangular matrix) to two commuting operators.

Commuting operators are simultaneously upper triangularizable Result 5.80

Suppose V is a finite-dimensional complex vector space and S, T are commuting operators on V. Then there is a basis of V with respect to which both S and T have upper-triangular matrices.

Proof. Let $n = \dim V$. We will use induction on n. The desired result holds if $n = 1$ because all 1-by-1 matrices are upper triangular. Now suppose $n > 1$ and the desired result holds for all complex vector spaces whose dimension is $n - 1$.

Let v_1 be any common eigenvector of S and T (using 5.78). Hence $Sv_1 \in \text{span}(v_1)$ and $Tv_1 \in \text{span}(v_1)$. Let W be a subspace of V such that

$$V = \text{span}(v_1) \oplus W;$$

see 2.33 for the existence of W. Define a linear map $P: V \to W$ by

$$P(av_1 + w) = w$$

for each $a \in \mathbb{C}$ and each $w \in W$. Define $\widehat{S}, \widehat{T} \in \text{End}(W)$ by

$$\widehat{S}w = P(Sw) \quad \text{and} \quad \widehat{T}w = P(Tw)$$

for each $w \in W$. To apply our induction hypothesis to $\widehat{S}$ and $\widehat{T}$, we must first show that these two operators on W commute. To do this, suppose $w \in W$. Then there exists $a \in \mathbb{C}$ such that

$$(\widehat{S}\widehat{T})w = \widehat{S}\big(P(Tw)\big) = \widehat{S}(Tw - av_1) = P\big(S(Tw - av_1)\big) = P\big((ST)w\big),$$

where the last equality holds because v_1 is an eigenvector of S and $Pv_1 = 0$. Similarly,

$$(\widehat{T}\widehat{S})w = P\big((TS)w\big).$$

Because the operators S and T commute, the last two displayed equations show that $(\widehat{S}\widehat{T})w = (\widehat{T}\widehat{S})w$. Hence $\widehat{S}$ and $\widehat{T}$ commute.

Thus we can use our induction hypothesis to state that there exists a basis $v_2, \ldots, v_n$ of W such that $\widehat{S}$ and $\widehat{T}$ both have upper-triangular matrices with respect to this basis. The list $v_1, \ldots, v_n$ is a basis of V.

If $k \in \{2, \ldots, n\}$, then there exist $a_k, b_k \in \mathbb{C}$ such that

$$Sv_k = a_k v_1 + \widehat{S}v_k \quad \text{and} \quad Tv_k = b_k v_1 + \widehat{T}v_k.$$

Because $\widehat{S}$ and $\widehat{T}$ have upper-triangular matrices with respect to $v_2, \ldots, v_n$, we know that $\widehat{S}v_k \in \text{span}(v_2, \ldots, v_k)$ and $\widehat{T}v_k \in \text{span}(v_2, \ldots, v_k)$. Hence the equations above imply that

$$Sv_k \in \text{span}(v_1, \ldots, v_k) \quad \text{and} \quad Tv_k \in \text{span}(v_1, \ldots, v_k).$$

Thus S and T have upper-triangular matrices with respect to $v_1, \ldots, v_n$, as desired. □

Exercise 9(b) extends the result above to more than two operators.

In general, it is not possible to determine the eigenvalues of the sum or product of two operators from the eigenvalues of the two operators. However, the next result shows that something nice happens when the two operators commute.

Eigenvalues of sum and product of commuting operators **Result 5.81**

Suppose V is a finite-dimensional complex vector space and S, T are commuting operators on V. Then

- every eigenvalue of $S + T$ is an eigenvalue of S plus an eigenvalue of T,
- every eigenvalue of ST is an eigenvalue of S times an eigenvalue of T.

Proof. There is a basis of V with respect to which both S and T have upper-triangular matrices (by 5.80). With respect to that basis,

$$\text{Mat}(S + T) = \text{Mat}(S) + \text{Mat}(T) \quad \text{and} \quad \text{Mat}(ST) = \text{Mat}(S)\,\text{Mat}(T),$$

as stated in 3.35 and 3.43.

The definition of matrix addition shows that each entry on the diagonal of $\mathrm{Mat}(S+T)$ equals the sum of the corresponding entries on the diagonals of $\mathrm{Mat}(S)$ and $\mathrm{Mat}(T)$. Similarly, because $\mathrm{Mat}(S)$ and $\mathrm{Mat}(T)$ are upper-triangular matrices, the definition of matrix multiplication shows that each entry on the diagonal of $\mathrm{Mat}(ST)$ equals the product of the corresponding entries on the diagonals of $\mathrm{Mat}(S)$ and $\mathrm{Mat}(T)$. Furthermore, $\mathrm{Mat}(S+T)$ and $\mathrm{Mat}(ST)$ are upper-triangular matrices (see Exercise 2 in Section 5C).

Every entry on the diagonal of $\mathrm{Mat}(S)$ is an eigenvalue of S, and every entry on the diagonal of $\mathrm{Mat}(T)$ is an eigenvalue of T (by 5.41). Every eigenvalue of $S+T$ is on the diagonal of $\mathrm{Mat}(S+T)$, and every eigenvalue of ST is on the diagonal of $\mathrm{Mat}(ST)$ (these assertions follow from 5.41). Putting all this together, we conclude that every eigenvalue of $S+T$ is an eigenvalue of S plus an eigenvalue of T, and every eigenvalue of ST is an eigenvalue of S times an eigenvalue of T. □

Exercises

5E.1 Give an example of two commuting operators S, T on $\mathbb{F}^4$ such that there is a subspace of $\mathbb{F}^4$ that is invariant under S but not under T and there is a subspace of $\mathbb{F}^4$ that is invariant under T but not under S.

5E.2 Suppose $\mathscr{E}$ is a subset of $\mathrm{End}(V)$ and every element of $\mathscr{E}$ is diagonalizable. Prove that there exists a basis of V with respect to which every element of $\mathscr{E}$ has a diagonal matrix if and only if every pair of elements of $\mathscr{E}$ commutes.

This exercise extends 5.76, which considers the case in which $\mathscr{E}$ contains only two elements. For this exercise, $\mathscr{E}$ may contain any number of elements, and $\mathscr{E}$ may even be an infinite set.

5E.3 Suppose $S, T \in \mathrm{End}(V)$ are such that $ST = TS$. Suppose $p \in \mathbb{F}[X]$.

(a) Prove that $\ker p(S)$ is invariant under T.

(b) Prove that $\mathrm{im}\, p(S)$ is invariant under T.

See 5.18 for the special case $S = T$.

5E.4 Prove or give a counterexample: If A is a diagonal matrix and B is an upper-triangular matrix of the same size as A, then A and B commute.

5E.5 Prove that a pair of operators on a finite-dimensional vector space commute if and only if their dual operators commute.

See 3.118 for the definition of the dual of an operator.

5E.6 Suppose that V is a nonzero finite-dimensional complex vector space and $S, T \in \mathrm{End}(V)$ commute. Prove that there exist $\alpha, \lambda \in \mathbb{C}$ such that

$$\mathrm{im}(S - \alpha I) + \mathrm{im}(T - \lambda I) \neq V.$$

5E.7 Suppose V is a complex vector space, $S \in \text{End}(V)$ is diagonalizable, and $T \in \text{End}(V)$ commutes with S. Prove that there is a basis of V such that S has a diagonal matrix with respect to this basis and T has an upper-triangular matrix with respect to this basis.

5E.8 Suppose $m = 3$ in Example 5.72 and D_x, D_y are the commuting partial differentiation operators on $\mathbb{R}^2[X]_{\le 3}$ from that example. Find a basis of $\mathbb{R}^2[X]_{\le 3}$ with respect to which D_x and D_y each have an upper-triangular matrix.

5E.9 Suppose V is a finite-dimensional nonzero complex vector space. Suppose that $\mathscr{E} \subseteq \text{End}(V)$ is such that S and T commute for all $S, T \in \mathscr{E}$.

(a) Prove that there is a vector in V that is an eigenvector for every element of $\mathscr{E}$.

(b) Prove that there is a basis of V with respect to which every element of $\mathscr{E}$ has an upper-triangular matrix.

This exercise extends 5.78 and 5.80, which consider the case in which $\mathscr{E}$ contains only two elements. For this exercise, $\mathscr{E}$ may contain any number of elements, and $\mathscr{E}$ may even be an infinite set.

5E.10 Give an example of two commuting operators S, T on a finite-dimensional real vector space such that $S + T$ has an eigenvalue that does not equal an eigenvalue of S plus an eigenvalue of T and ST has an eigenvalue that does not equal an eigenvalue of S times an eigenvalue of T.

This exercise shows that 5.81 does not hold on real vector spaces.

Part 3



Inner Products

CHAPTER 6	Inner Product Spaces	184
CHAPTER 7	Operators on Inner Product Spaces	230

6 Inner Product Spaces

In making the definition of a vector space, we generalized the linear structure (addition and scalar multiplication) of $\mathbb{R}^2$ and $\mathbb{R}^3$. We ignored geometric features such as the notions of length and angle. These ideas are embedded in the concept of inner products, which we will investigate in this chapter.

Every inner product induces a norm, which you can think of as a length. This norm satisfies key properties such as the Pythagorean theorem, the triangle inequality, the parallelogram equality, and the Cauchy–Schwarz inequality.

The notion of perpendicular vectors in Euclidean geometry gets renamed to orthogonal vectors in the context of an inner product space. We will see that orthonormal bases are tremendously useful in inner product spaces. The Gram–Schmidt procedure constructs such bases. This chapter will conclude by putting together these tools to solve minimization problems.

Standing assumptions for this chapter

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

6A Inner Products and Norms

Inner Products

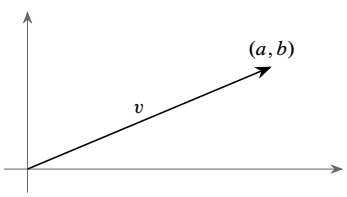
To motivate the concept of inner product, think of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ as arrows with initial point at the origin. The length of a vector v in $\mathbb{R}^2$ or $\mathbb{R}^3$ is called the *norm* of v and is denoted by $\|v\|$. Thus for $v = (a, b) \in \mathbb{R}^2$, we have

$$\|v\| = \sqrt{a^2 + b^2}.$$

Similarly, if $v = (a, b, c) \in \mathbb{R}^3$, then $\|v\| = \sqrt{a^2 + b^2 + c^2}$.

Even though we cannot draw pictures in higher dimensions, the generalization to $\mathbb{R}^n$ is easy: we define the norm of $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ by

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$



This vector v has norm $\sqrt{a^2 + b^2}$.

Definition

The norm is not linear on $\mathbb{R}^n$. To inject linearity into the discussion, we introduce the dot product.

Dot product Definition 6.1

For $x, y \in \mathbb{R}^n$, the *dot product* of x and y, denoted by $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

The dot product of two vectors in $\mathbb{R}^n$ is a number, not a vector. Notice that $x \cdot x = \|x\|^2$ for all $x \in \mathbb{R}^n$. Furthermore, the dot product on $\mathbb{R}^n$ has the following properties.

If we think of a vector as a point instead of as an arrow, then $\|x\|$ should be interpreted to mean the distance from the origin to the point x.

- $x \cdot x \geq 0$ for all $x \in \mathbb{R}^n$.
- $x \cdot x = 0$ if and only if $x = 0$.
- For $y \in \mathbb{R}^n$ fixed, the map from $\mathbb{R}^n$ to $\mathbb{R}$ that sends $x \in \mathbb{R}^n$ to $x \cdot y$ is linear.
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^n$.

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 $a, b \in \mathbb{R}$, then

- the absolute value of λ, denoted by $|\lambda|$, is defined by $|\lambda| = \sqrt{a^2 + b^2}$;
- the complex conjugate of λ, denoted by $\overline{\lambda}$, is defined by $\overline{\lambda} = a - bi$;
- $|\lambda|^2 = \lambda\overline{\lambda}$.

See Chapter 4 for the definitions and the basic properties of the absolute value and complex conjugate.

For $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$, 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 $\mathbb{R}^n$. The equation above thus suggests that the inner product of the vector $w = (w_1, \ldots, w_n) \in \mathbb{C}^n$ 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. Thus 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.

One comment about the notation used in the next definition:

- For $\lambda \in \mathbb{C}$, the notation $\lambda \geq 0$ means λ is real and nonnegative.

Definition 6.2

Inner product

An *inner product* on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbb{F}$ 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 $\lambda \in \mathbb{F}$ and all $u, v \in V$.

conjugate symmetry
$\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$.

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 $u, v \in V$.

Most mathematicians define inner products as above, but many physicists use a definition that requires homogeneity in the second slot instead of the first slot.

Example 6.3 **Inner products**

(a) The *Euclidean inner product* on $\mathbb{F}^n$ is defined by

$$\big\langle (w_1, \dots, w_n), (z_1, \dots, z_n)\big\rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n}$$

for all $(w_1, \dots, w_n), (z_1, \dots, z_n) \in \mathbb{F}^n$.

(b) If $c_1, \dots, c_n$ are positive numbers, then an inner product can be defined on $\mathbb{F}^n$ by

$$\big\langle (w_1, \dots, w_n), (z_1, \dots, z_n)\big\rangle = c_1 w_1\overline{z_1} + \cdots + c_n w_n\overline{z_n}$$

for all $(w_1, \dots, w_n), (z_1, \dots, z_n) \in \mathbb{F}^n$.

(c) An inner product can be defined on the vector space of continuous real-valued functions on the interval $[-1, 1]$ by

$$\langle f, g\rangle = \int_{-1}^{1} fg$$

for all f, g continuous real-valued functions on $[-1, 1]$.

Example 6.3
continued

(d) An inner product can be defined on $\mathbb{R}[X]$ by

$$\langle p, q\rangle = p(0)q(0) + \int_{-1}^{1} p'q'$$

for all $p, q \in \mathbb{R}[X]$.

(e) An inner product can be defined on $\mathbb{R}[X]$ by

$$\langle p, q\rangle = \int_{0}^{\infty} p(x)q(x)e^{-x}\,dx$$

for all $p, q \in \mathbb{R}[X]$.

Inner product space Definition 6.4

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 $\mathbb{F}^n$ with the Euclidean inner product given by (a) in the example above. When $\mathbb{F}^n$ 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 and W are inner product spaces, we make the following assumption.

V, W Notation 6.5

For the rest of this chapter and the next chapter, V and W denote inner product spaces over $\mathbb{F}$.

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 clear from the context (or is the Euclidean inner product if the vector space is $\mathbb{F}^n$).

Basic properties of an inner product Result 6.6

(a) For each fixed $v \in V$, the function that takes $u \in V$ to $\langle u, v\rangle$ is a linear map from V to $\mathbb{F}$.

(b) $\langle 0, v\rangle = 0$ for every $v \in V$.

(c) $\langle v, 0\rangle = 0$ for every $v \in V$.

(d) $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$ for all $u, v, w \in V$.

(e) $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$ for all $\lambda \in \mathbb{F}$ and all $u, v \in V$.

Proof.

(a) For $v \in V$, the linearity of $u \mapsto \langle u, v\rangle$ follows from the conditions of additivity and homogeneity in the first slot in the definition of an inner product.

(b) Every linear map takes 0 to 0. Thus (b) follows from (a).

(c) If $v \in V$, then the conjugate symmetry property in the definition of an inner product and (b) show that $\langle v, 0\rangle = \overline{\langle 0, v\rangle} = \overline{0} = 0$.

(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 $\lambda \in \mathbb{F}$ and $u, v \in V$. Then

$$\begin{aligned}\langle u, \lambda v\rangle &= \overline{\langle \lambda v, u\rangle}\\ &= \overline{\lambda \langle v, u\rangle}\\ &= \overline{\lambda}\, \overline{\langle v, u\rangle}\\ &= \overline{\lambda}\langle u, v\rangle.\end{aligned}$$

□

Norms

Our motivation for defining inner products came initially from the norms of vectors on $\mathbb{R}^2$ and $\mathbb{R}^3$. Now we see that each inner product determines a norm.

Norm, $\|v\|$ Definition 6.7

For $v \in V$, the *norm* of v, denoted by $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Example 6.8 **Norms**

(a) If $(z_1, \ldots, z_n) \in \mathbb{F}^n$ (with the Euclidean inner product), then

$$\|(z_1, \ldots, z_n)\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

(b) For f in the vector space of continuous real-valued functions on $[-1, 1]$ and

Example 6.8
continued

with inner product given as in 6.3(c), we have

$$\|f\| = \sqrt{\int_{-1}^{1} f^2}.$$

Basic properties of the norm — Result 6.9

Suppose $v \in V$.

(a) $\|v\| = 0$ if and only if $v = 0$.

(b) $\|\lambda v\| = |\lambda| \, \|v\|$ for all $\lambda \in \mathbb{F}$.

Proof.

(a) The desired result holds because $\langle v, v\rangle = 0$ if and only if $v = 0$.

(b) Suppose $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} \|\lambda v\|^2 &= \langle \lambda v, \lambda v\rangle \\ &= \lambda \langle v, \lambda v\rangle \\ &= \lambda \overline{\lambda} \langle v, v\rangle \\ &= |\lambda|^2 \|v\|^2. \end{aligned}$$

Taking square roots now gives the desired equality. □

The proof of (b) in the result above illustrates a general principle: working with norms squared is usually easier than working directly with norms.

Now we come to a crucial definition.

Orthogonal — Definition 6.10

Two vectors $u, v \in V$ are called *orthogonal* if $\langle u, v\rangle = 0$.

The word **orthogonal** comes from the Greek word **orthogonios**, which means right-angled.

In the definition above, the order of the two vectors does not matter, because $\langle u, v\rangle = 0$ if and only if $\langle v, u\rangle = 0$. Instead of saying u and v are orthogonal, sometimes we say u is orthogonal to v.

Exercise 15 asks you to prove that if u, v are nonzero vectors in $\mathbb{R}^2$, 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 nonzero vectors in $\mathbb{R}^2$ are orthogonal (with respect to the 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.

Orthogonality and 0 **Result 6.11**

(a) 0 is orthogonal to every vector in V.

(b) 0 is the only vector in V that is orthogonal to itself.

Proof.

(a) Recall that 6.6(b) states that $\langle 0, v\rangle = 0$ for every $v \in V$.

(b) If $v \in V$ and $\langle v, v\rangle = 0$, then $v = 0$ (by definition of inner product). □

For the special case $V = \mathbb{R}^2$, the next theorem was known over 3,500 years ago in Babylonia and then rediscovered and proved over 2,500 years ago in Greece. Of course, the proof below is not the original proof.

Pythagorean theorem **Result 6.12**

Suppose $u, v \in V$. If u and v are orthogonal, then

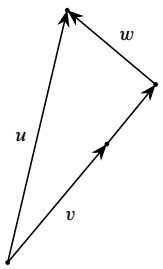
$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof. Suppose $\langle u, v\rangle = 0$. Then

$$\begin{aligned}\|u + v\|^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle \\ &= \|u\|^2 + \|v\|^2.\end{aligned}$$

□

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 picture here.



An orthogonal decomposition: u expressed as a scalar multiple of v plus a vector orthogonal to v.

To discover how to write u as a scalar multiple of v plus a vector orthogonal to v, let $c \in \mathbb{F}$ denote a scalar. Then

$$u = cv + (u - cv).$$

Thus we need to choose c so that v is orthogonal to $(u - cv)$. Hence 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 displayed above explicitly writes u as a scalar multiple of v plus a vector orthogonal to v. Thus we have proved the following key result.

An orthogonal decomposition **Result 6.13**

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

$$u = cv + w \quad \text{and} \quad \langle w, v\rangle = 0.$$

The orthogonal decomposition 6.13 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.

Cauchy–Schwarz inequality **Result 6.14**

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 6.13, where w is orthogonal to v. By the Pythagorean theorem,

$$\begin{aligned} \|u\|^2 &= \left\| \frac{\langle u, v\rangle}{\|v\|^2} v \right\|^2 + \|w\|^2 \\ &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\ &\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}. \end{aligned} \tag{6.16}$$

Multiplying both sides of this inequality by $\|v\|^2$ and then taking square roots gives the desired inequality.

The proof in the paragraph above shows that the Cauchy–Schwarz inequality is an equality if and only if 6.15 is an equality. This happens if and only if $w = 0$. But $w = 0$ if and only if u is a multiple of v (see 6.13). 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). □

Augustin-Louis Cauchy (1789–1857) proved 6.17(a) in 1821. In 1859, Cauchy's student Viktor Bunyakovsky (1804–1889) proved integral inequalities like the one in 6.17(b). A few decades later, similar discoveries by Hermann Schwarz (1843–1921) attracted more attention and led to the name of this inequality.

Example 6.17 **Cauchy–Schwarz inequality**

(a) If $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$, then

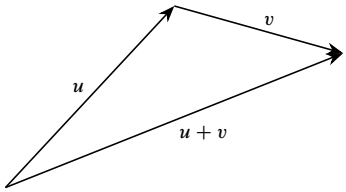
$$(x_1 y_1 + \cdots + x_n y_n)^2 \le (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2),$$

as follows from applying the Cauchy–Schwarz inequality to the vectors $(x_1, \dots, x_n), (y_1, \dots, y_n) \in \mathbb{R}^n$, using the usual Euclidean inner product.

(b) If f, g are continuous real-valued functions on $[-1, 1]$, then

$$\left| \int_{-1}^{1} fg \right|^2 \le \left(\int_{-1}^{1} f^2 \right) \left(\int_{-1}^{1} g^2 \right),$$

as follows from applying the Cauchy–Schwarz inequality to Example 6.3(c).



In this triangle, the length of $u + v$ is less than the length of u plus the length of v.

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 polygonal path between two points is a single line segment (a polygonal path consists of line segments).

Result 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 real multiple of the other.

Proof. We have

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v \rangle \\
&= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle \\
&= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \overline{\langle u, v \rangle} \\
&= \|u\|^2 + \|v\|^2 + 2 \operatorname{Re} \langle u, v \rangle \\
&\le \|u\|^2 + \|v\|^2 + 2 |\langle u, v \rangle| \qquad (6.20) \\
&\le \|u\|^2 + \|v\|^2 + 2 \|u\| \, \|v\| \qquad (6.22) \\
&= (\|u\| + \|v\|)^2,
\end{aligned}$$

where 6.22 follows from the Cauchy–Schwarz inequality (6.14). Taking square roots of both sides of the inequality above gives the desired inequality.

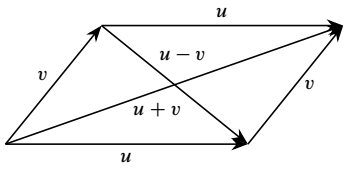
The proof above shows that the triangle inequality is an equality if and only if we have equality in 6.19 and 6.22. Thus we have equality in the triangle inequality if and only if

$$\langle u, v \rangle = \|u\| \, \|v\|. \qquad (6.24)$$

If one of u, v is a nonnegative real multiple of the other, then 6.23 holds. Conversely, suppose 6.23 holds. Then the condition for equality in the Cauchy–Schwarz inequality (6.14) implies that one of u, v is a scalar multiple of the other. This scalar must be a nonnegative real number, by 6.23, completing the proof. □

For the reverse triangle inequality, see Exercise 20.

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. Note that the proof here is more straightforward than the usual proof in Euclidean geometry.



The diagonals of this parallelogram are $u + v$ and $u - v$.

Parallelogram equality **Result 6.25**

Suppose $u, v \in V$. Then

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Proof. We have

$$\begin{aligned}\|u + v\|^2 + \|u - v\|^2 &= \langle u + v, u + v\rangle + \langle u - v, u - v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\ &\quad + \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\ &= 2(\|u\|^2 + \|v\|^2),\end{aligned}$$

as desired. □

Exercises

6A.1 Prove or give a counterexample: If $v_1, \dots, v_m \in V$, then

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\langle v_j, v_k\rangle \geq 0.$$

6A.2 Suppose $S \in \text{End}(V)$. Define $\langle \cdot, \cdot \rangle_1$ by

$$\langle u, v\rangle_1 = \langle Su, Sv\rangle$$

for all $u, v \in V$. Show that $\langle \cdot, \cdot \rangle_1$ is an inner product on V if and only if S is injective.

6A.3

(a) Show that the function taking an ordered pair $\big((x_1, x_2), (y_1, y_2)\big)$ of elements of $\mathbb{R}^2$ to $|x_1 y_1| + |x_2 y_2|$ is not an inner product on $\mathbb{R}^2$.

(b) Show that the function taking an ordered pair $\big((x_1, x_2, x_3), (y_1, y_2, y_3)\big)$ of elements of $\mathbb{R}^3$ to $x_1 y_1 + x_3 y_3$ is not an inner product on $\mathbb{R}^3$.

6A.4 Suppose $T \in \text{End}(V)$ is such that $\|Tv\| \le \|v\|$ for every $v \in V$. Prove that $T - \sqrt{2}\, I$ is injective.

6A.5 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 (b) to show that the diagonals of a rhombus are perpendicular to each other.

6A.6 Suppose $u, v \in V$. Prove that $\langle u, v\rangle = 0 \iff \|u\| \le \|u + av\|$ for all $a \in \mathbb{F}$.

6A.7 Suppose $u, v \in V$. Prove that $\|au + bv\| = \|bu + av\|$ for all $a, b \in \mathbb{R}$ if and only if $\|u\| = \|v\|$.

6A.8 Suppose $a, b, c, x, y \in \mathbb{R}$ and $a^2 + b^2 + c^2 + x^2 + y^2 \le 1$. Prove that $a + b + c + 4x + 9y \le 10$.

6A.9 Suppose $u, v \in V$ and $\|u\| = \|v\| = 1$ and $\langle u, v\rangle = 1$. Prove that $u = v$.

6A.10 Suppose $u, v \in V$ and $\|u\| \le 1$ and $\|v\| \le 1$. Prove that

$$\sqrt{1 - \|u\|^2}\,\sqrt{1 - \|v\|^2} \le 1 - |\langle u, v\rangle|.$$

6A.11 Find vectors $u, v \in \mathbb{R}^2$ such that u is a scalar multiple of $(1, 3)$, v is orthogonal to $(1, 3)$, and $(1, 2) = u + v$.

6A.12 Suppose a, b, c, d are positive numbers.

(a) Prove that $(a+b+c+d)\Big(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\Big) \ge 16$.

(b) For which positive numbers a, b, c, d is the inequality above an equality?

6A.13 Show that the square of an average is less than or equal to the average of the squares. More precisely, show that if $a_1, \ldots, a_n \in \mathbb{R}$, then the square of the average of $a_1, \ldots, a_n$ is less than or equal to the average of $a_1^2, \ldots, a_n^2$.

6A.14 Suppose $v \in V$ and $v \ne 0$. Prove that $v/\|v\|$ is the unique closest element on the unit sphere of V to v. More precisely, prove that if $u \in V$ and $\|u\| = 1$, then

$$\left\| v - \frac{v}{\|v\|} \right\| \le \|v - u\|,$$

with equality only if $u = v/\|v\|$.

6A.15 Suppose u, v are nonzero vectors in $\mathbb{R}^2$. 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: Use the law of cosines on the triangle formed by u, v, and $u - v$.

6A.16 The angle between two vectors (thought of as arrows with initial point at the origin) in $\mathbb{R}^2$ or $\mathbb{R}^3$ can be defined geometrically. However, geometry is not as clear in $\mathbb{R}^n$ for $n > 3$. Thus the angle between two nonzero vectors $x, y \in \mathbb{R}^n$ is defined to be

$$\arccos\frac{\langle x, y\rangle}{\|x\|\,\|y\|},$$

where the motivation for this definition comes from Exercise 15. Explain why the Cauchy–Schwarz inequality is needed to show that this definition makes sense.

6A.17 Prove that

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 \le \left(\sum_{k=1}^{n} k a_k^2\right)\left(\sum_{k=1}^{n} \frac{b_k^2}{k}\right)$$

for all real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$.

6A.18

(a) Suppose $f : [1, \infty) \to [0, \infty)$ is continuous. Show that

$$\left(\int_1^\infty f\right)^2 \le \int_1^\infty x^2 \big(f(x)\big)^2\, dx.$$

(b) For which continuous functions $f : [1, \infty) \to [0, \infty)$ is the inequality in (a) an equality with both sides finite?

6A.19 Suppose $v_1, \ldots, v_n$ is a basis of V and $T \in \text{End}(V)$. Prove that if λ is an eigenvalue of T, then

$$|\lambda|^2 \le \sum_{j=1}^{n}\sum_{k=1}^{n} |\text{Mat}(T)_{j,k}|^2,$$

where $\text{Mat}(T)_{j,k}$ denotes the entry in row j, column k of the matrix of T with respect to the basis $v_1, \ldots, v_n$.

6A.20 Prove that if $u, v \in V$, then $\big|\,\|u\| - \|v\|\,\big| \le \|u - v\|$.

*The inequality above is called the **reverse triangle inequality**. For the reverse triangle inequality when $V = \mathbb{C}$, see Exercise 2 in Chapter 4.*

6A.21 Suppose $u, v \in V$ are such that

$$\|u\| = 3, \qquad \|u + v\| = 4, \qquad \|u - v\| = 6.$$

What number does $\|v\|$ equal?

6A.22 Show that if $u, v \in V$, then

$$\|u + v\| \, \|u - v\| \le \|u\|^2 + \|v\|^2.$$

6A.23 Suppose $v_1, \dots, v_m \in V$ are such that $\|v_k\| \le 1$ for each $k = 1, \dots, m$. Show that there exist $a_1, \dots, a_m \in \{1, -1\}$ such that

$$\|a_1 v_1 + \cdots + a_m v_m\| \le \sqrt{m}.$$

6A.24 Prove or give a counterexample: If $\|\cdot\|$ is the norm associated with an inner product on $\mathbb{R}^2$, then there exists $(x, y) \in \mathbb{R}^2$ such that $\|(x, y)\| \ne \max\{|x|, |y|\}$.

6A.25 Suppose $p > 0$. Prove that there is an inner product on $\mathbb{R}^2$ such that the associated norm is given by

$$\|(x, y)\| = (|x|^p + |y|^p)^{1/p}$$

for all $(x, y) \in \mathbb{R}^2$ if and only if $p = 2$.

6A.26 Suppose V is a real inner product space. Prove that

$$\langle u, v \rangle = \frac{\|u + v\|^2 - \|u - v\|^2}{4}$$

for all $u, v \in V$.

6A.27 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 $u, v \in V$.

6A.28 A norm on a vector space U is a function

$$\|\cdot\| \colon U \to [0, \infty)$$

such that $\|u\| = 0$ if and only if $u = 0$, $\|\alpha u\| = |\alpha| \|u\|$ for all $\alpha \in \mathbb{F}$ and all $u \in U$, and $\|u + v\| \le \|u\| + \|v\|$ 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 $\|\cdot\|$ is a norm on U satisfying the parallelogram equality, then there is an inner product $\langle \cdot, \cdot \rangle$ on U such that $\|u\| = \langle u, u \rangle^{1/2}$ for all $u \in U$).

6A.29 Suppose $V_1, \dots, V_m$ are inner product spaces. Show that the equation

$$\big\langle (u_1, \dots, u_m), (v_1, \dots, v_m) \big\rangle = \langle u_1, v_1 \rangle + \cdots + \langle u_m, v_m \rangle$$

defines an inner product on $V_1 \times \cdots \times V_m$.

In the expression above on the right, for each $k = 1, \dots, m$, the inner product $\langle u_k, v_k \rangle$ denotes the inner product on V_k. Each of the spaces $V_1, \dots, V_m$ may have a different inner product, even though the same notation is used here.

6A.30 Suppose V is a real inner product space. For $u, v, w, x \in V$, define

$$\langle u + iv, w + ix \rangle_{\mathbb{C}} = \langle u, w \rangle + \langle v, x \rangle + (\langle v, w \rangle - \langle u, x \rangle)i.$$

(a) Show that $\langle \cdot, \cdot \rangle_{\mathbb{C}}$ makes $V_{\mathbb{C}}$ into a complex inner product space.

(b) Show that if $u, v \in V$, then

$$\langle u, v \rangle_{\mathbb{C}} = \langle u, v \rangle \quad \text{and} \quad \|u + iv\|_{\mathbb{C}}^2 = \|u\|^2 + \|v\|^2.$$

See Exercise 8 in Section 1B for the definition of the complexification $V_{\mathbb{C}}$.

6A.31 Suppose $u, v, w \in V$. Prove that

$$\big\|w - (1/2)(u + v)\big\|^2 = \frac{\|w - u\|^2 + \|w - v\|^2}{2} - \frac{\|u - v\|^2}{4}.$$

6A.32 Suppose that E is a subset of V with the property that $u, v \in E$ implies $(1/2)(u + v) \in E$. Let $w \in V$. Show that there is at most one point in E that is closest to w. In other words, show that there is at most one $u \in E$ such that

$$\|w - u\| \le \|w - x\|$$

for all $x \in E$.

6A.33 Suppose f, g are differentiable functions from $\mathbb{R}$ to $\mathbb{R}^n$.

(a) Show that

$$\big\langle f(t), g(t) \big\rangle' = \big\langle f'(t), g(t) \big\rangle + \big\langle f(t), g'(t) \big\rangle.$$

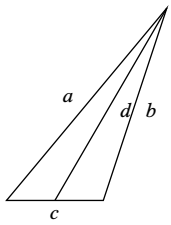
(b) Suppose c is a positive number and $\|f(t)\| = c$ for every $t \in \mathbb{R}$. Show that $\big\langle f'(t), f(t) \big\rangle = 0$ for every $t \in \mathbb{R}$.

(c) Interpret the result in (b) geometrically in terms of the tangent vector to a curve lying on a sphere in $\mathbb{R}^n$ centered at the origin.

A function $f : \mathbb{R} \to \mathbb{R}^n$ is called differentiable if there exist differentiable functions $f_1, \dots, f_n$ from $\mathbb{R}$ to $\mathbb{R}$ such that $f(t) = \big(f_1(t), \dots, f_n(t)\big)$ for each $t \in \mathbb{R}$. Furthermore, for each $t \in \mathbb{R}$, the derivative $f'(t) \in \mathbb{R}^n$ is defined by $f'(t) = \big(f_1'(t), \dots, f_n'(t)\big)$.

6A.34 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

$$a^2 + b^2 = 1/2c^2 + 2d^2.$$



6A.35 Fix a positive integer n. The *Laplacian* Δp of a twice differentiable real-valued function p on $\mathbb{R}^n$ is the function on $\mathbb{R}^n$ 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 $\mathbb{R}^n$ is a linear combination (with coefficients in $\mathbb{R}$) of functions of the form $x_1^{m_1} \cdots x_n^{m_n}$, where $m_1, \ldots, m_n$ are nonnegative integers.

Suppose q is a polynomial on $\mathbb{R}^n$. Prove that there exists a harmonic polynomial p on $\mathbb{R}^n$ such that $p(x) = q(x)$ for every $x \in \mathbb{R}^n$ with $\|x\| = 1$.

The only fact about harmonic functions that you need for this exercise is that if p is a harmonic function on $\mathbb{R}^n$ and $p(x) = 0$ for all $x \in \mathbb{R}^n$ 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 $\mathbb{R}^n$ 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.

In realms of numbers, where the secrets lie,
A noble truth emerges from the deep,
Cauchy and Schwarz, their wisdom they apply,
An inequality for all to keep.
Two vectors, by this bond, are intertwined,
As inner products weave a gilded thread,
Their magnitude, by providence, confined,
A bound to which their destiny is wed.
Though shadows fall, and twilight dims the day,
This inequality will stand the test,
To guide us in our quest, to light the way,
And in its truth, our understanding rest.
So sing, ye muses, of this noble feat,
Cauchy–Schwarz, the bound that none can beat.

—written by ChatGPT with input *Shakespearean sonnet on Cauchy–Schwarz inequality*

6B Orthonormal Bases

Orthonormal Lists and the Gram–Schmidt Procedure

Orthonormal — Definition 6.26

- 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 $e_1, \dots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases}$$

for all $j, k \in \{1, \dots, m\}$.

Example 6.27 **Orthonormal lists**

(a) The standard basis of $\mathbb{F}^n$ is an orthonormal list.

(b) $\big(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}\big), \big(-1/\sqrt{2}, 1/\sqrt{2}, 0\big)$ is an orthonormal list in $\mathbb{F}^3$.

(c) $\big(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}\big), \big(-1/\sqrt{2}, 1/\sqrt{2}, 0\big), \big(1/\sqrt{6}, 1/\sqrt{6}, -2/\sqrt{6}\big)$ is an orthonormal list in $\mathbb{F}^3$.

Example 6.27
continued

(d) Suppose n is a positive integer. Then, as Exercise 4 asks you to verify,

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\sin nx}{\sqrt{\pi}}$$

is an orthonormal list of vectors in $C[-\pi, \pi]$, the vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} fg.$$

The orthonormal list above is often used for modeling periodic phenomena, such as tides.

(e) Suppose we make $\mathbb{R}[X]_{\le 2}$ into an inner product space using the inner product given by

$$\langle p, q \rangle = \int_{-1}^{1} pq$$

for all $p, q \in \mathbb{R}[X]_{\le 2}$. The standard basis $1, x, x^2$ of $\mathbb{R}[X]_{\le 2}$ is not an orthonormal list because the vectors in that list do not have norm 1. Dividing each vector by its norm gives the list $1/\sqrt{2}$, $\sqrt{3/2}\, x$, $\sqrt{5/2}\, x^2$, in which each vector has norm 1, and the second vector is orthogonal to the first and third vectors. However, the first and third vectors are not orthogonal. Thus this is not an orthonormal list. Soon we will see how to construct an orthonormal list from the standard basis $1, x, x^2$ (see Example 6.38).

Orthonormal lists are particularly easy to work with, as illustrated by the next result.

Norm of an orthonormal linear combination Result 6.28

Suppose $e_1, \dots, e_m$ 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 $a_1, \dots, a_m \in \mathbb{F}$.

Proof. Because each e_k has norm 1, this follows from repeated applications of the Pythagorean theorem (6.12). □

The result above has the following important corollary.

Orthonormal lists are linearly independent Result 6.29

Every orthonormal list of vectors is linearly independent.

Proof. Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in V and $a_1, \dots, a_m \in \mathbb{F}$ are such that

$$a_1 e_1 + \cdots + a_m e_m = 0.$$

Then $|a_1|^2 + \cdots + |a_m|^2 = 0$ (by 6.28), which means that all the a_k's are 0. Thus $e_1, \ldots, e_m$ is linearly independent. □

Now we come to an important inequality.

Bessel's inequality — Result 6.30

Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in V. If $v \in V$ then

$$|\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2 \le \|v\|^2.$$

Proof. Suppose $v \in V$. Then

$$v = \underbrace{\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m}_{u} + \underbrace{v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m}_{w}.$$

Let u and w be defined as in the equation above. If $k \in \{1, \ldots, m\}$, then $\langle w, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle\langle e_k, e_k\rangle = 0$. This implies that $\langle w, u\rangle = 0$. The Pythagorean theorem now implies that

$$\begin{aligned}\|v\|^2 &= \|u\|^2 + \|w\|^2 \\ &\ge \|u\|^2 \\ &= |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2,\end{aligned}$$

where the last line comes from 6.28. □

The next definition introduces one of the most useful concepts in the study of inner product spaces.

Orthonormal basis — Definition 6.31

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 $\mathbb{F}^n$.

Orthonormal lists of the right length are orthonormal bases — Result 6.32

Suppose V is finite-dimensional. Then every orthonormal list of vectors in V of length $\dim V$ is an orthonormal basis of V.

Proof. By 6.29, every orthonormal list of vectors in V is linearly independent. Thus every such list of the right length is a basis—see 2.38. □

Example 6.33 **An orthonormal basis of $\mathbb{F}^4$**

As mentioned above, the standard basis is an orthonormal basis of $\mathbb{F}^4$. We now

Example 6.33
continued

show that

$$(1/2, 1/2, 1/2, 1/2), (1/2, 1/2, -1/2, -1/2), (1/2, -1/2, -1/2, 1/2), (-1/2, 1/2, -1/2, 1/2)$$

is also an orthonormal basis of $\mathbb{F}^4$.

We have

$$\big\|(1/2, 1/2, 1/2, 1/2)\big\| = \sqrt{1/2^2 + 1/2^2 + 1/2^2 + 1/2^2} = 1.$$

Similarly, the other three vectors in the list above also have norm 1.

Note that

$$\big\langle (1/2, 1/2, 1/2, 1/2), (1/2, 1/2, -1/2, -1/2) \big\rangle = 0.$$

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 $\mathbb{F}^4$, this list is an orthonormal basis of $\mathbb{F}^4$ (by 6.32).

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 $a_1, \dots, a_n \in \mathbb{F}$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Computing the numbers $a_1, \dots, a_n$ that satisfy the equation above can be a long computation for an arbitrary basis of V. The next result shows, however, that this is easy for an orthonormal basis—just take $a_k = \langle v, e_k \rangle$.

The formula below for $\|v\|$ is called Parseval's identity. It was published in 1799 in the context of Fourier series.

Notice how the next result makes each inner product space of dimension n behave like $\mathbb{F}^n$, with the role of the coordinates of a vector in $\mathbb{F}^n$ played by $\langle v, e_1 \rangle, \dots, \langle v, e_n \rangle$.

Writing a vector as a linear combination of an orthonormal basis Result 6.34

Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $u, v \in V$. Then

(a) $v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$;

(b) $\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$;

(c) $\langle u, v \rangle = \langle u, e_1 \rangle \overline{\langle v, e_1 \rangle} + \cdots + \langle u, e_n \rangle \overline{\langle v, e_n \rangle}$.

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_k gives $\langle v, e_k \rangle = a_k$. Thus (a) holds.

Now (b) follows immediately from (a) and 6.28.

Take the inner product of u with each side of (a) and then get (c) by using conjugate linearity [6.6(d) and 6.6(e)] in the second slot of the inner product. □

Example 6.35 **Finding coefficients for a linear combination**

Suppose we want to write the vector $(1, 2, 4, 7) \in \mathbb{F}^4$ as a linear combination of the orthonormal basis

$$(1/2, 1/2, 1/2, 1/2), (1/2, 1/2, -1/2, -1/2), (1/2, -1/2, -1/2, 1/2), (-1/2, 1/2, -1/2, 1/2)$$

of $\mathbb{F}^4$ from Example 6.33. Instead of solving a system of four linear equations in four unknowns, as typically would be required if we were working with a nonorthonormal basis, we simply evaluate four inner products and use 6.34(a), getting that $(1, 2, 4, 7)$ equals

$$7(1/2, 1/2, 1/2, 1/2) - 4(1/2, 1/2, -1/2, -1/2) + (1/2, -1/2, -1/2, 1/2) + 2(-1/2, 1/2, -1/2, 1/2).$$

Jørgen Gram (1850–1916) and Erhard Schmidt (1876–1959) popularized this algorithm that constructs orthonormal lists.

Now that we understand the usefulness of orthonormal bases, how do we go about finding them? For example, does $\mathbb{R}[X]_{\le m}$ with inner product as in 6.3(c) have an orthonormal basis? The next result will lead to answers to these questions.

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.

Result 6.36 **Gram–Schmidt procedure**

Suppose $v_1, \dots, v_m$ is a linearly independent list of vectors in V. Let $f_1 = v_1$. For $k = 2, \dots, m$, define f_k inductively by

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}.$$

For each $k = 1, \dots, m$, let $e_k = \dfrac{f_k}{\|f_k\|}$. Then $e_1, \dots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$$

for each $k = 1, \dots, m$.

Proof. We will show by induction on k that the desired conclusion holds. To get started with $k = 1$, note that because $e_1 = f_1/\|f_1\|$, we have $\|e_1\| = 1$; also, $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ because e_1 is a nonzero multiple of v_1.

Suppose $1 < k \le m$ and the list $e_1, \dots, e_{k-1}$ generated by 6.36 is an orthonormal list such that

$$\operatorname{span}(v_1, \dots, v_{k-1}) = \operatorname{span}(e_1, \dots, e_{k-1}). \tag{6.37}$$

Because $v_1, \dots, v_m$ is linearly independent, we have $v_k \notin \operatorname{span}(v_1, \dots, v_{k-1})$. Thus $v_k \notin \operatorname{span}(e_1, \dots, e_{k-1}) = \operatorname{span}(f_1, \dots, f_{k-1})$, which implies that $f_k \ne 0$. Hence we are not dividing by 0 in the definition of e_k given in 6.36. Dividing a vector by its norm produces a new vector with norm 1; thus $\|e_k\| = 1$.

Let $j \in \{1, \dots, k-1\}$. Then

$$\begin{aligned}\langle e_k, e_j \rangle &= \frac{1}{\|f_k\| \, \|f_j\|} \langle f_k, f_j \rangle \\ &= \frac{1}{\|f_k\| \, \|f_j\|} \left\langle v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}, f_j \right\rangle \\ &= \frac{1}{\|f_k\| \, \|f_j\|} (\langle v_k, f_j \rangle - \langle v_k, f_j \rangle) \\ &= 0.\end{aligned}$$

Thus $e_1, \dots, e_k$ is an orthonormal list.

From the definition of e_k given in 6.36, we see that $v_k \in \operatorname{span}(e_1, \dots, e_k)$. Combining this information with 6.37 shows that

$$\operatorname{span}(v_1, \dots, v_k) \subseteq \operatorname{span}(e_1, \dots, e_k).$$

Both lists above are linearly independent (the v's by hypothesis, and the e's by orthonormality and 6.29). Thus both subspaces above have dimension k, and hence they are equal, completing the induction step and thus completing the proof. □

Example 6.38 **An orthonormal basis of $\mathbb{R}[X]_{\le 2}$**

Suppose we make $\mathbb{R}[X]_{\le 2}$ into an inner product space using the inner product given by

$$\langle p, q \rangle = \int_{-1}^{1} pq$$

for all $p, q \in \mathbb{R}[X]_{\le 2}$. We know that $1, x, x^2$ is a basis of $\mathbb{R}[X]_{\le 2}$, but it is not an orthonormal basis. We will find an orthonormal basis of $\mathbb{R}[X]_{\le 2}$ by applying the Gram–Schmidt procedure with $v_1 = 1$, $v_2 = x$, and $v_3 = x^2$.

To get started, take $f_1 = v_1 = 1$. Thus $\|f_1\|^2 = \int_{-1}^{1} 1 = 2$. Hence the formula in 6.36 tells us that

$$f_2 = v_2 - \frac{\langle v_2, f_1 \rangle}{\|f_1\|^2} f_1 = x - \frac{\langle x, 1 \rangle}{\|f_1\|^2} = x,$$

where the last equality holds because $\langle x, 1 \rangle = \int_{-1}^{1} t \, dt = 0$.

The formula above for f_2 implies that $\|f_2\|^2 = \int_{-1}^{1} t^2 \, dt = 2/3$. Now the formula in 6.36 tells us that

$$f_3 = v_3 - \frac{\langle v_3, f_1 \rangle}{\|f_1\|^2} f_1 - \frac{\langle v_3, f_2 \rangle}{\|f_2\|^2} f_2 = x^2 - 1/2\langle x^2, 1 \rangle - 3/2\langle x^2, x \rangle x = x^2 - 1/3.$$

The formula above for f_3 implies that

$$\|f_3\|^2 = \int_{-1}^{1} (t^2 - 1/3)^2 \, dt = \int_{-1}^{1} (t^4 - 2/3t^2 + 1/9) \, dt = 8/45.$$

Example 6.38 *continued*

Now dividing each of f_1, f_2, f_3 by its norm gives us the orthonormal list

$$\sqrt{1/2},\ \sqrt{3/2}\,x,\ \sqrt{45/8}(x^2 - 1/3).$$

The orthonormal list above has length three, which is the dimension of $\mathbb{R}[X]_{\le 2}$. Hence this orthonormal list is an orthonormal basis of $\mathbb{R}[X]_{\le 2}$ [by 6.32].

Now we can answer the question about the existence of orthonormal bases.

Existence of orthonormal basis — Result 6.39

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 (6.36) to it, producing an orthonormal list of length $\dim V$. By 6.32, 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.

Every orthonormal list extends to an orthonormal basis — Result 6.40

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 6.29). Hence this list can be extended to a basis $e_1, \dots, e_m, v_1, \dots, v_n$ of V (see 2.32). Now apply the Gram–Schmidt procedure (6.36) to $e_1, \dots, e_m, v_1, \dots, v_n$, producing an orthonormal list

$$e_1, \dots, e_m, f_1, \dots, f_n;$$

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 6.32. □

Recall that a matrix is called upper triangular if it 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 gave a necessary and sufficient condition for an operator to have an upper-triangular matrix with respect to some basis (see 5.44). 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 condition for an operator to have an upper-triangular matrix with respect to some orthonormal basis is the same as the condition to have an upper-triangular matrix with respect to an arbitrary basis.

Upper-triangular matrix with respect to some orthonormal basis — Result 6.41

Suppose V is finite-dimensional and $T \in \text{End}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V if and only if the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Proof. Suppose T has an upper-triangular matrix with respect to some basis $v_1, \ldots, v_n$ of V. Thus $\text{span}(v_1, \ldots, v_k)$ is invariant under T for each $k = 1, \ldots, n$ (see 5.39).

Apply the Gram–Schmidt procedure to $v_1, \ldots, v_n$, producing an orthonormal basis $e_1, \ldots, e_n$ of V. Because

$$\text{span}(e_1, \ldots, e_k) = \text{span}(v_1, \ldots, v_k)$$

for each k (see 6.36), we conclude that $\text{span}(e_1, \ldots, e_k)$ is invariant under T for each $k = 1, \ldots, n$. Thus, by 5.39, T has an upper-triangular matrix with respect to the orthonormal basis $e_1, \ldots, e_n$. Now use 5.44 to complete the proof. □

Issai Schur (1875–1941) published a proof of the next result in 1909.

For complex vector spaces, the next result is an important application of the result above. See Exercise 20 for a version of Schur's theorem that applies simultaneously to more than one operator.

Schur's theorem — Result 6.42

Every operator on a finite-dimensional complex inner product space has an upper-triangular matrix with respect to some orthonormal basis.

Proof. The desired result follows from the second version of the fundamental theorem of algebra (4.13) and 6.41. □

Linear Functionals on Inner Product Spaces

Because linear maps into the scalar field $\mathbb{F}$ play a special role, we defined a special name for them and their vector space in Section 3F. Those definitions are repeated below in case you skipped Section 3F.

Linear functional, dual space, V' — Definition 6.43

- A *linear functional* on V is a linear map from V to $\mathbb{F}$.
- The *dual space* of V, denoted by V', is the vector space of all linear functionals on V. In other words, $V' = \text{Hom}(V, \mathbb{F})$.

Example 6.44 **Linear functional on $\mathbb{F}^3$**

The function $\varphi: \mathbb{F}^3 \to \mathbb{F}$ defined by

$$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

is a linear functional on $\mathbb{F}^3$. We could write this linear functional in the form

$$\varphi(z) = \langle z, w \rangle$$

for every $z \in \mathbb{F}^3$, where $w = (2, -5, 1)$.

Example 6.45 **Linear functional on $\mathbb{R}[X]_{\le 5}$**

The function $\varphi: \mathbb{R}[X]_{\le 5} \to \mathbb{R}$ defined by

$$\varphi(p) = \int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\, dt$$

is a linear functional on $\mathbb{R}[X]_{\le 5}$.

If $v \in V$, then the map that sends u to $\langle u, v \rangle$ is a linear functional on V. The next result states that every linear functional on V is of this form. For example, we can take $v = (2, -5, 1)$ in Example 6.44.

Suppose we make the vector space $\mathbb{R}[X]_{\le 5}$ into an inner product space by defining $\langle p, q \rangle = \int_{-1}^{1} pq$. Let φ be as in Example 6.45. It is not obvious that there exists $q \in \mathbb{R}[X]_{\le 5}$ such that

$$\int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\, dt = \langle p, q \rangle$$

for every $p \in \mathbb{R}[X]_{\le 5}$ [we cannot take $q(t) = \cos(\pi t)$ because that choice of q is not an element of $\mathbb{R}[X]_{\le 5}$]. The next result tells us the somewhat surprising result that there indeed exists a polynomial $q \in \mathbb{R}[X]_{\le 5}$ such that the equation above holds for all $p \in \mathbb{R}[X]_{\le 5}$.

The next result is named in honor of Frigyes Riesz (1880–1956), who proved several theorems early in the twentieth century that look very much like the result below.

Result 6.46

Riesz representation theorem
Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $v \in V$ such that

$$\varphi(u) = \langle u, v \rangle$$

for every $u \in V$.

Proof. First we show that there exists a vector $v \in V$ such that $\varphi(u) = \langle u, v \rangle$ for every

$u \in V$. Let $e_1, \dots, e_n$ be an orthonormal basis of V. Then

$$\begin{aligned}\varphi(u) &= \varphi(\langle u, e_1\rangle e_1 + \cdots + \langle u, e_n\rangle e_n)\\ &= \langle u, e_1\rangle \varphi(e_1) + \cdots + \langle u, e_n\rangle \varphi(e_n)\\ &= \big\langle u, \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n\big\rangle\end{aligned}$$

for every $u \in V$, where the first equality comes from 6.34(a). Thus setting

$$v = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n, \tag{6.47}$$

we have $\varphi(u) = \langle u, v\rangle$ for every $u \in V$, as desired.

Now we prove that only one vector $v \in V$ has the desired behavior. Suppose $v_1, v_2 \in V$ are such that

$$\varphi(u) = \langle u, v_1\rangle = \langle u, v_2\rangle$$

for every $u \in V$. Then

$$0 = \langle u, v_1\rangle - \langle u, v_2\rangle = \langle u, v_1 - v_2\rangle$$

for every $u \in V$. Taking $u = v_1 - v_2$ shows that $v_1 - v_2 = 0$. Thus $v_1 = v_2$, completing the proof of the uniqueness part of the result. □

Example 6.48 **Computation illustrating Riesz representation theorem**

Suppose we want to find a polynomial $q \in \mathbb{R}[X]_{\le 2}$ such that

$$\int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\, dt = \int_{-1}^{1} pq \tag{6.49}$$

for every polynomial $p \in \mathbb{R}[X]_{\le 2}$. To do this, we make $\mathbb{R}[X]_{\le 2}$ into an inner product space by defining $\langle p, q\rangle$ to be the right side of the equation above for $p, q \in \mathbb{R}[X]_{\le 2}$. Note that the left side of the equation above does not equal the inner product in $\mathbb{R}[X]_{\le 2}$ of p and the function $t \mapsto \cos(\pi t)$ because this last function is not a polynomial.

Define a linear functional φ on $\mathbb{R}[X]_{\le 2}$ by letting

$$\varphi(p) = \int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\, dt$$

for each $p \in \mathbb{R}[X]_{\le 2}$. Now use the orthonormal basis from Example 6.38 and apply formula 6.47 from the proof of the Riesz representation theorem to see that if $p \in \mathbb{R}[X]_{\le 2}$, then $\varphi(p) = \langle p, q\rangle$, where

$$\begin{aligned}q(x) = {} & \left(\int_{-1}^{1} \sqrt{1/2}\cos(\pi t)\, dt\right)\sqrt{1/2} + \left(\int_{-1}^{1} \sqrt{3/2}\, t\cos(\pi t)\, dt\right)\sqrt{3/2}\, x\\ & + \left(\int_{-1}^{1} \sqrt{45/8}(t^2 - 1/3)\cos(\pi t)\, dt\right)\sqrt{45/8}(x^2 - 1/3).\end{aligned}$$

Example 6.48
continued

A bit of calculus applied to the equation above shows that

$$q(x) = \tfrac{15}{2\pi^2}(1 - 3x^2).$$

The same procedure shows that if we want to find $q \in \mathbb{R}[X]_{\le 5}$ such that 6.49 holds for all $p \in \mathbb{R}[X]_{\le 5}$, then we should take

$$q(x) = \tfrac{105}{8\pi^4}\Big((27 - 2\pi^2) + (24\pi^2 - 270)x^2 + (315 - 30\pi^2)x^4\Big).$$

Suppose V is finite-dimensional and φ a linear functional on V. Then 6.47 gives a formula for the vector v that satisfies

$$\varphi(u) = \langle u, v\rangle$$

for all $u \in V$. Specifically, we have

$$v = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n.$$

The right side of the equation above seems to depend on the orthonormal basis $e_1, \ldots, e_n$ as well as on φ. However, 6.46 tells us that v is uniquely determined by φ. Thus the right side of the equation above is the same regardless of which orthonormal basis $e_1, \ldots, e_n$ of V is chosen.

For two additional different proofs of the Riesz representation theorem, see 6.62 and also Exercise 13 in Section 6C.

Exercises

6B.1 Suppose $e_1, \ldots, e_m$ is a list of vectors in V such that

$$\|a_1e_1 + \cdots + a_me_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \ldots, a_m \in \mathbb{F}$. Show that $e_1, \ldots, e_m$ is an orthonormal list.

This exercise provides a converse to 6.28.

6B.2

(a) Suppose $\theta \in \mathbb{R}$. Show that both

$$(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta) \quad \text{and} \quad (\cos\theta, \sin\theta), (\sin\theta, -\cos\theta)$$

are orthonormal bases of $\mathbb{R}^2$.

(b) Show that each orthonormal basis of $\mathbb{R}^2$ is of the form given by one of the two possibilities in (a).

6B.3 Suppose $e_1, \dots, e_m$ is an orthonormal list in V and $v \in V$. Prove that

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2 \iff v \in \operatorname{span}(e_1, \dots, e_m).$$

6B.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

$$\langle f, g\rangle = \int_{-\pi}^{\pi} fg.$$

Hint: The following formulas should help.

$$(\sin x)(\cos y) = \frac{\sin(x-y) + \sin(x+y)}{2}$$
$$(\sin x)(\sin y) = \frac{\cos(x-y) - \cos(x+y)}{2}$$
$$(\cos x)(\cos y) = \frac{\cos(x-y) + \cos(x+y)}{2}$$

6B.5 Suppose $f : [-\pi, \pi] \to \mathbb{R}$ is continuous. For each nonnegative integer k, define

$$a_k = \tfrac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x)\cos(kx)\,dx \quad \text{and} \quad b_k = \tfrac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(x)\sin(kx)\,dx.$$

Prove that

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty}(a_k^2 + b_k^2) \le \int_{-\pi}^{\pi} f^2.$$

The inequality above is actually an equality for all continuous functions $f : [-\pi, \pi] \to \mathbb{R}$. However, proving that this inequality is an equality involves Fourier series techniques beyond the scope of this book.

6B.6 Suppose $e_1, \dots, e_n$ is an orthonormal basis of V.

(a) Prove that if $v_1, \dots, v_n$ are vectors in V such that

$$\|e_k - v_k\| < \frac{1}{\sqrt{n}}$$

for each k, then $v_1, \dots, v_n$ is a basis of V.

(b) Show that there exist $v_1, \dots, v_n \in V$ such that

$$\|e_k - v_k\| \le \frac{1}{\sqrt{n}}$$

for each k, but $v_1, \dots, v_n$ is not linearly independent.

This exercise states in (a) *that an appropriately small perturbation of an orthonormal basis is a basis. Then* (b) *shows that the number $1/\sqrt{n}$ on the right side of the inequality in* (a) *cannot be higher.*

6B.7 Suppose $T \in \text{End}(\mathbb{R}^3)$ has an upper-triangular matrix with respect to the basis $(1,0,0)$, $(1,1,1)$, $(1,1,2)$. Find an orthonormal basis of $\mathbb{R}^3$ with respect to which T has an upper-triangular matrix.

6B.8 Make $\mathbb{R}[X]_{\le 2}$ into an inner product space by defining $\langle p, q\rangle = \int_0^1 pq$ for all $p, q \in \mathbb{R}[X]_{\le 2}$.

(a) Apply the Gram–Schmidt procedure to the basis $1, x, x^2$ to produce an orthonormal basis of $\mathbb{R}[X]_{\le 2}$.

(b) The differentiation operator (the operator that takes p to p') on $\mathbb{R}[X]_{\le 2}$ has an upper-triangular matrix with respect to the basis $1, x, x^2$, which is not an orthonormal basis. Find the matrix of the differentiation operator on $\mathbb{R}[X]_{\le 2}$ with respect to the orthonormal basis produced in (a) and verify that this matrix is upper triangular, as expected from the proof of 6.41.

6B.9 Suppose $e_1, \dots, e_m$ is the result of applying the Gram–Schmidt procedure to a linearly independent list $v_1, \dots, v_m$ in V. Prove that $\langle v_k, e_k\rangle > 0$ for each $k = 1, \dots, m$.

6B.10 Suppose $v_1, \dots, v_m$ is a linearly independent list in V. Explain why the orthonormal list produced by the formulas of the Gram–Schmidt procedure (6.36) is the only orthonormal list $e_1, \dots, e_m$ in V such that $\langle v_k, e_k\rangle > 0$ and $\text{span}(v_1, \dots, v_k) = \text{span}(e_1, \dots, e_k)$ for each $k = 1, \dots, m$.

The result in this exercise is used in the proof of 7.58.

6B.11 Find a polynomial $q \in \mathbb{R}[X]_{\le 2}$ such that $p(1/2) = \int_0^1 pq$ for every $p \in \mathbb{R}[X]_{\le 2}$.

6B.12 Find a polynomial $q \in \mathbb{R}[X]_{\le 2}$ such that

$$\int_0^1 p(x)\cos(\pi x)\,dx = \int_0^1 pq$$

for every $p \in \mathbb{R}[X]_{\le 2}$.

6B.13 Show that a list $v_1, \dots, v_m$ of vectors in V is linearly dependent if and only if the Gram–Schmidt formula in 6.36 produces $f_k = 0$ for some $k \in \{1, \dots, m\}$.

This exercise gives an alternative to Gaussian elimination techniques for determining whether a list of vectors in an inner product space is linearly dependent.

6B.14 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

$$\text{span}(v_1, \dots, v_k) = \text{span}(e_1, \dots, e_k)$$

for all $k \in \{1, \dots, m\}$.

6B.15 Suppose $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ are inner products on V such that $\langle u, v \rangle_1 = 0$ if and only if $\langle u, v \rangle_2 = 0$. Prove that there is a positive number c such that $\langle u, v \rangle_1 = c \langle u, v \rangle_2$ for every $u, v \in V$.

This exercise shows that if two inner products have the same pairs of orthogonal vectors, then each of the inner products is a scalar multiple of the other inner product.

6B.16 Suppose V is finite-dimensional. Suppose $\langle \cdot, \cdot \rangle_1$, $\langle \cdot, \cdot \rangle_2$ are inner products on V with corresponding norms $\|\cdot\|_1$ and $\|\cdot\|_2$. Prove that there exists a positive number c such that $\|v\|_1 \le c\|v\|_2$ for every $v \in V$.

6B.17 Suppose $\mathbb{F} = \mathbb{C}$ and V is finite-dimensional. Prove that if T is an operator on V such that 1 is the only eigenvalue of T and $\|Tv\| \le \|v\|$ for all $v \in V$, then T is the identity operator.

6B.18 Suppose $u_1, \dots, u_m$ is a linearly independent list in V. Show that there exists $v \in V$ such that $\langle u_k, v \rangle = 1$ for all $k \in \{1, \dots, m\}$.

6B.19 Suppose $v_1, \dots, v_n$ is a basis of V. Prove that there exists a basis $u_1, \dots, u_n$ of V such that

$$\langle v_j, u_k \rangle = \begin{cases} 0 & \text{if } j \ne k, \\ 1 & \text{if } j = k. \end{cases}$$

6B.20 Suppose $\mathbb{F} = \mathbb{C}$, V is finite-dimensional, and $\mathscr{E} \subseteq \operatorname{End}(V)$ is such that

$$ST = TS$$

for all $S, T \in \mathscr{E}$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathscr{E}$ has an upper-triangular matrix.

This exercise strengthens Exercise 9(b) in Section 5E (in the context of inner product spaces) by asserting that the basis in that exercise can be chosen to be orthonormal.

6B.21 Suppose $\mathbb{F} = \mathbb{C}$, V is finite-dimensional, $T \in \operatorname{End}(V)$, and all eigenvalues of T have absolute value less than 1. Let $\epsilon > 0$. Prove that there exists a positive integer m such that $\|T^m v\| \le \epsilon \|v\|$ for every $v \in V$.

6B.22 Suppose $C[-1, 1]$ is the vector space of continuous real-valued functions on the interval $[-1, 1]$ with inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} fg$$

for all $f, g \in C[-1, 1]$. Let φ be the linear functional on $C[-1, 1]$ defined by $\varphi(f) = f(0)$. Show that there does not exist $g \in C[-1, 1]$ such that

$$\varphi(f) = \langle f, g \rangle$$

for every $f \in C[-1, 1]$.

This exercise shows that the Riesz representation theorem (6.46) does not hold on infinite-dimensional vector spaces without additional hypotheses on V and φ.

6B.23 For all $u, v \in V$, 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).

This exercise requires familiarity with metric spaces.

Remark (Orthogonality at the Supreme Court). 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.)

6C Orthogonal Complements and Minimization Problems

Orthogonal Complements

Orthogonal complement, $U^\perp$ — Definition 6.50

If U is a subset of V, then the *orthogonal complement* of U, denoted by $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle u, v \rangle = 0 \text{ for every } u \in U\}.$$

The orthogonal complement $U^\perp$ depends on V as well as on U. However, the inner product space V should always be clear from the context and thus it can be omitted from the notation.

Example 6.51 **Orthogonal complements**

- If $V = \mathbb{R}^3$ and U is the subset of V consisting of the single point $(2, 3, 5)$, then $U^\perp$ is the plane $\{(x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = 0\}$.
- If $V = \mathbb{R}^3$ and U is the plane $\{(x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = 0\}$, then $U^\perp$ is the line $\{(2t, 3t, 5t) : t \in \mathbb{R}\}$.
- More generally, if U is a plane in $\mathbb{R}^3$ containing the origin, then $U^\perp$ is the line containing the origin that is perpendicular to U.
- If U is a line in $\mathbb{R}^3$ containing the origin, then $U^\perp$ is the plane containing the origin that is perpendicular to U.
- If $V = \mathbb{F}^5$ and $U = \{(a, b, 0, 0, 0) \in \mathbb{F}^5 : a, b \in \mathbb{F}\}$, then
$$U^\perp = \{(0, 0, x, y, z) \in \mathbb{F}^5 : x, y, z \in \mathbb{F}\}.$$
- If $e_1, \dots, e_m, f_1, \dots, f_n$ is an orthonormal basis of V, then
$$\big(\operatorname{span}(e_1, \dots, e_m)\big)^\perp = \operatorname{span}(f_1, \dots, f_n).$$

We begin with some straightforward consequences of the definition.

Properties of orthogonal complement **Result 6.52**

(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 $U \cap U^\perp \subseteq \{0\}$.

(e) If G and H are subsets of V and $G \subseteq H$, then $H^\perp \subseteq G^\perp$.

Proof.

(a) Suppose U is a subset of V. Then $\langle u, 0\rangle = 0$ for every $u \in U$; thus $0 \in U^\perp$.

Suppose $v, w \in U^\perp$. If $u \in U$, then
$$\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle = 0 + 0 = 0.$$
Thus $v + w \in U^\perp$, which shows that $U^\perp$ is closed under addition.

Similarly, suppose $\lambda \in \mathbb{F}$ and $v \in U^\perp$. If $u \in U$, then
$$\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle = \overline{\lambda} \cdot 0 = 0.$$
Thus $\lambda v \in U^\perp$, which shows that $U^\perp$ is closed under scalar multiplication. Thus $U^\perp$ is a subspace of V.

(b) Suppose that $v \in V$. Then $\langle 0, v\rangle = 0$, which implies that $v \in \{0\}^\perp$. Thus $\{0\}^\perp = V$.

(c) Suppose that $v \in V^\perp$. Then $\langle v, v\rangle = 0$, which implies that $v = 0$. Thus $V^\perp = \{0\}$.

(d) Suppose U is a subset of V and $u \in U \cap U^\perp$. Then $\langle u, u\rangle = 0$, which implies that $u = 0$. Thus $U \cap U^\perp \subseteq \{0\}$.

(e) Suppose G and H are subsets of V and $G \subseteq H$. Suppose $v \in H^\perp$. Then $\langle u, v\rangle = 0$ for every $u \in H$, which implies that $\langle u, v\rangle = 0$ for every $u \in G$. Hence $v \in G^\perp$. Thus $H^\perp \subseteq G^\perp$. □

Recall that if U and 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 1.41). Furthermore, this happens if and only if $V = U + W$ and $U \cap W = \{0\}$ (see 1.46).

The next result shows that every finite-dimensional subspace of V leads to a natural direct sum decomposition of V. See Exercise 16 for an example showing that the result below can fail without the hypothesis that the subspace U is finite-dimensional.

Direct sum of a subspace and its orthogonal complement **Result 6.53**

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^\perp.$$

Proof. First we will show that

$$V = U + U^\perp.$$

To do this, suppose that $v \in V$. Let $e_1, \dots, e_m$ be an orthonormal basis of U. We want to write v as the sum of a vector in U and a vector orthogonal to U.

We have

$$v = \underbrace{\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m}_{u} + \underbrace{v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m}_{w}. \tag{6.54}$$

Let u and w be defined as in the equation above (as was done in the proof of 6.30). Because each $e_k \in U$, we see that $u \in U$. Because $e_1, \dots, e_m$ is an orthonormal list, for each $k = 1, \dots, m$ we have

$$\begin{aligned}\langle w, e_k\rangle &= \langle v, e_k\rangle - \langle v, e_k\rangle \\ &= 0.\end{aligned}$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \dots, e_m)$, which shows that $w \in U^\perp$. Hence we have written $v = u + w$, where $u \in U$ and $w \in U^\perp$, completing the proof that $V = U + U^\perp$.

From 6.52(d), we know that $U \cap U^\perp = \{0\}$. Now equation $V = U + U^\perp$ implies that $V = U \oplus U^\perp$ (see 1.46). □

Now we can see how to compute $\dim U^\perp$ from $\dim U$.

Dimension of orthogonal complement — Result 6.55

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 6.53 and 3.94. □

The next result is an important consequence of 6.53.

Orthogonal complement of the orthogonal complement — Result 6.56

Suppose U is a finite-dimensional subspace of V. Then

$$U = (U^\perp)^\perp.$$

Proof. First we will show that

$$U \subseteq (U^\perp)^\perp. \tag{6.57}$$

To do this, suppose $u \in U$. Then $\langle u, w\rangle = 0$ for every $w \in U^\perp$ (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 6.57.

To prove the inclusion in the other direction, suppose $v \in (U^\perp)^\perp$. By 6.53, 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 6.57), we have $v - u \in (U^\perp)^\perp$. Thus $v - u \in U^\perp \cap (U^\perp)^\perp$, which implies that $v - u = 0$ [by 6.52(d)], which implies that $v = u$, which implies that $v \in U$. Thus $(U^\perp)^\perp \subseteq U$, which along with 6.57 completes the proof. □

Exercise 16(a) shows that the result below is not true without the hypothesis that U is finite-dimensional.

Suppose U is a subspace of V and we want to show that $U = V$. In some situations, the easiest way to do this is to show that the only vector orthogonal to U is 0, and then use the result below. For example, the result below is useful for Exercise 4.

$U^\perp = \{0\} \iff U = V$ (for U a finite-dimensional subspace of V) — Result 6.58

Suppose U is a finite-dimensional subspace of V. Then

$$U^\perp = \{0\} \iff U = V.$$

Proof. First suppose $U^\perp = \{0\}$. Then by 6.56, $U = (U^\perp)^\perp = \{0\}^\perp = V$, as desired.

Conversely, if $U = V$, then $U^\perp = V^\perp = \{0\}$ by 6.52(c). □

We now define an operator P_U for each finite-dimensional subspace U of V.

Orthogonal projection, P_U — Definition 6.59

Suppose U is a finite-dimensional subspace of V. The *orthogonal projection* of V onto U is the operator $P_U \in \text{End}(V)$ defined as follows: For each $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then let $P_U v = u$.

The direct sum decomposition $V = U \oplus U^\perp$ given by 6.53 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. See the figure that accompanies the proof of 6.65 for the picture describing $P_U v$ that you should keep in mind.

Example 6.60 **Orthogonal projection onto one-dimensional subspace**

Suppose $u \in V$ with $u \neq 0$ and U is the one-dimensional subspace of V defined by $U = \operatorname{span}(u)$.

If $v \in V$, then

$$v = \frac{\langle v, u\rangle}{\|u\|^2} u + \left(v - \frac{\langle v, u\rangle}{\|u\|^2} u\right),$$

where the first term on the right is in $\operatorname{span}(u)$ (and thus is in U) and the second term on the right is orthogonal to u (and thus is in $U^\perp$). Thus $P_U v$ equals the first term on the right. In other words, we have the formula

$$P_U v = \frac{\langle v, u\rangle}{\|u\|^2} u$$

for every $v \in V$.

The formula above becomes $P_U u = u$ if $v = u$ and becomes $P_U v = 0$ if $v \in \{u\}^\perp$. These equations are special cases of (b) and (c) in the next result.

Result 6.61 **Properties of orthogonal projection P_U**

Suppose U is a finite-dimensional subspace of V. Then

(a) $P_U \in \operatorname{End}(V)$;

(b) $P_U u = u$ for every $u \in U$;

(c) $P_U w = 0$ for every $w \in U^\perp$;

(d) $\operatorname{im} P_U = U$;

(e) $\ker P_U = U^\perp$;

(f) $v - P_U v \in U^\perp$ for every $v \in V$;

(g) $P_U^2 = P_U$;

(h) $\|P_U v\| \leq \|v\|$ for every $v \in V$;

(i) if $e_1, \ldots, e_m$ is an orthonormal basis of U and $v \in V$, then

$$P_U v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m.$$

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 $\lambda \in \mathbb{F}$ and $v \in V$. Write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then $\lambda v = \lambda u + \lambda w$ 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 $\operatorname{im} P_U \subseteq U$. Furthermore, (b) implies that $U \subseteq \operatorname{im} P_U$. Thus $\operatorname{im} P_U = U$.

(e) The inclusion $U^\perp \subseteq \ker P_U$ follows from (c). To prove the inclusion in the other direction, note that if $v \in \ker P_U$ then the decomposition given by 6.53 must be $v = 0 + v$, where $0 \in U$ and $v \in U^\perp$. Thus $\ker P_U \subseteq U^\perp$.

(f) If $v \in V$ and $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$v - P_U v = v - u = w \in U^\perp.$$

(g) If $v \in V$ and $v = u + w$ 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 $v \in V$ and $v = u + w$ 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 6.54 in the proof of 6.53. □

In the previous section we proved the Riesz representation theorem (6.46), whose key part states that every linear functional on a finite-dimensional inner product space is given by taking the inner product with some fixed vector. Seeing a different proof often provides new insight. Thus we now give a new proof of the key part of the Riesz representation theorem using orthogonal complements instead of orthonormal bases as in our previous proof.

The restatement below of the Riesz representation theorem provides an identification of V with V'. We will prove only the "onto" part of the result below because the more routine "one-to-one" part of the result can be proved as in 6.46.

Intuition behind this new proof: If $\varphi \in V'$, $v \in V$, and $\varphi(u) = \langle u, v \rangle$ for all $u \in V$, then $v \in (\ker \varphi)^\perp$. However, $(\ker \varphi)^\perp$ is a one-dimensional subspace of V (except for the trivial case in which $\varphi = 0$), as follows from 6.55 and 3.21. Thus we can obtain v by choosing any nonzero element of $(\ker \varphi)^\perp$ and then multiplying by an appropriate scalar, as is done in the proof below.

Riesz representation theorem, revisited **Result 6.62**

Suppose V is finite-dimensional. For each $v \in V$, define $\varphi_v \in V'$ by

$$\varphi_v(u) = \langle u, v \rangle$$

for each $u \in V$. Then $v \mapsto \varphi_v$ is a one-to-one function from V onto V'.

Caution: The function $v \mapsto \varphi_v$ is a linear mapping from V to V' if $\mathbb{F} = \mathbb{R}$. However, this function is not linear if $\mathbb{F} = \mathbb{C}$ because $\varphi_{\lambda v} = \overline{\lambda}\varphi_v$ if $\lambda \in \mathbb{C}$.

Proof. To show that $v \mapsto \varphi_v$ is surjective, suppose $\varphi \in V'$. If $\varphi = 0$, then $\varphi = \varphi_0$. Thus assume $\varphi \neq 0$. Hence $\ker \varphi \neq V$, which implies that $(\ker \varphi)^\perp \neq \{0\}$ (by 6.53 with $U = \ker \varphi$).

Let $w \in (\ker \varphi)^\perp$ be such that $w \neq 0$. Let

$$v = \frac{\overline{\varphi(w)}}{\|w\|^2} w. \tag{6.63}$$

Then $v \in (\ker \varphi)^\perp$. Also, $v \neq 0$ (because $w \notin \ker \varphi$).

Taking the norm of both sides of 6.63 gives

$$\|v\| = \frac{|\varphi(w)|}{\|w\|}. \tag{6.64}$$

Applying φ to both sides of 6.63 and then using 6.64, we have

$$\varphi(v) = \frac{|\varphi(w)|^2}{\|w\|^2} = \|v\|^2.$$

Now suppose $u \in V$. Using the equation above, we have

$$u = \left(u - \frac{\varphi(u)}{\varphi(v)} v\right) + \frac{\varphi(u)}{\|v\|^2} v.$$

The first term in parentheses above is in $\ker \varphi$ and hence is orthogonal to v. Thus taking the inner product of both sides of the equation above with v shows that

$$\langle u, v \rangle = \frac{\varphi(u)}{\|v\|^2} \langle v, v \rangle = \varphi(u).$$

Thus $\varphi = \varphi_v$, showing that $v \mapsto \varphi_v$ is surjective, as desired. □

See Exercise 13 for yet another proof of the Riesz representation theorem.

Minimization Problems

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 result shows that $u = P_U v$ is the unique solution of this minimization problem.

The remarkable simplicity of the solution to this minimization problem has led to many important applications of inner product spaces outside of pure mathematics.

Minimizing distance to a subspace **Result 6.65**

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \le \|v - u\|.$$

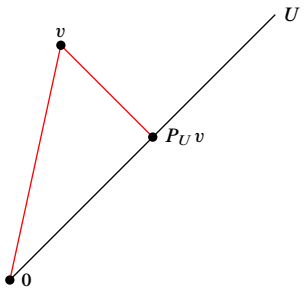
Furthermore, the inequality above is an equality if and only if $u = P_U v$.

Proof. We have

$$\begin{aligned} \|v - P_U v\|^2 &\le \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \big\|(v - P_U v) + (P_U v - u)\big\|^2 \\ &= \|v - u\|^2, \end{aligned} \tag{6.66}$$

where the first line above holds because $0 \le \|P_U v - u\|^2$, the second line above comes from the Pythagorean theorem [which applies because $v - P_U v \in U^\perp$ by 6.61(f), and $P_U v - u \in U$], and the third line above holds by simple algebra. Taking square roots gives the desired inequality.

The inequality proved above is an equality if and only if 6.66 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 6.61(i) to compute explicit solutions to minimization problems, as in the following example.

Example 6.67 **Using linear algebra to approximate the sine function**

Suppose we want to find a polynomial u with real coefficients and of degree at most 5 that approximates the sine function 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.

Let $C[-\pi, \pi]$ denote the real inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} fg. \tag{6.68}$$

Let $v \in C[-\pi, \pi]$ be the function defined by $v(x) = \sin x$. Let U denote the subspace of $C[-\pi, \pi]$ consisting of the polynomials with real coefficients and of 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.

To compute the solution to our approximation problem, first apply the Gram–Schmidt procedure (using the inner product given by 6.68) 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.

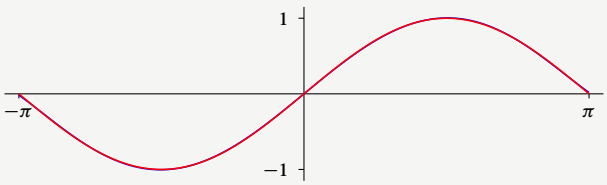
Example 6.67
continued

Then, again using the inner product given by 6.68, compute $P_U v$ using 6.61(i) (with $m = 6$). Doing this computation shows that $P_U v$ is the function u defined by

$$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5, \tag{6.69}$$

where the π's that appear in the exact answer have been replaced with a good decimal approximation. By 6.65, the polynomial u above is the best approximation to the sine function 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 the sine function and our approximation u given by 6.69 over the interval $[-\pi, \pi]$.



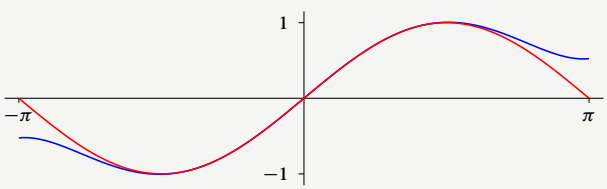
Graphs on $[-\pi, \pi]$ of the sine function (red) and its best fifth degree polynomial approximation u (blue) from 6.69.

Our approximation 6.69 is so accurate that the two graphs are almost identical—our eyes may see only one graph! Here the red graph is placed almost exactly over the blue graph. If you are viewing this on an electronic device, enlarge the picture above by 400% near π or $-\pi$ to see a small gap between the two graphs.

Another well-known approximation to the sine function by a polynomial of degree 5 is given by the Taylor polynomial p defined by

$$p(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}. \tag{6.70}$$

To see how good this approximation is, the next picture shows the graphs of both the sine function and the Taylor polynomial p over the interval $[-\pi, \pi]$.



Graphs on $[-\pi, \pi]$ of the sine function (red) and the Taylor polynomial (blue) from 6.70.

The Taylor polynomial of degree 5 is an excellent approximation to $\sin x$ for x near 0. But the picture above shows that for $|x| > 2$, the Taylor polynomial is

Example 6.67 *continued*

not so accurate, especially compared to 6.69. For example, taking $x = 3$, our approximation 6.69 estimates $\sin 3$ with an error of approximately 0.001, but the Taylor polynomial 6.70 estimates $\sin 3$ with an error of approximately 0.4. Thus at $x = 3$, the error in the Taylor polynomial is hundreds of times larger than the error given by 6.69. Linear algebra has helped us discover an approximation to the sine function that improves upon what we learned in calculus!

Pseudoinverse

Suppose $T \in \text{Hom}(V, W)$ and $w \in W$. Consider the problem of finding $v \in V$ such that

$$Tv = w.$$

For example, if $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, then the equation above could represent a system of m linear equations in n unknowns $v_1, \dots, v_n$, where $v = (v_1, \dots, v_n)$.

If T is invertible, then the unique solution to the equation above is $v = T^{-1}w$. However, if T is not invertible, then for some $w \in W$ there may not exist any solutions of the equation above, and for some $w \in W$ there may exist infinitely many solutions of the equation above.

If T is not invertible, then we can still try to do as well as possible with the equation above. For example, if the equation above has no solutions, then instead of solving the equation $Tv - w = 0$, we can try to find $v \in V$ such that $\|Tv - w\|$ is as small as possible. As another example, if the equation above has infinitely many solutions $v \in V$, then among all those solutions we can try to find one such that $\|v\|$ is as small as possible.

The pseudoinverse will provide the tool to solve the equation above as well as possible, even when T is not invertible. We need the next result to define the pseudoinverse.

In the next two proofs, we will use without further comment the result that if V is finite-dimensional and $T \in \text{Hom}(V, W)$, then $\ker T$, $(\ker T)^\perp$, and $\text{im}\, T$ are all finite-dimensional.

Result 6.71 **Restriction of a linear map to obtain a one-to-one and onto map**

Suppose V is finite-dimensional and $T \in \text{Hom}(V, W)$. Then $T|_{(\ker T)^\perp}$ is an injective map of $(\ker T)^\perp$ onto $\text{im}\, T$.

Proof. Suppose that $v \in (\ker T)^\perp$ and $T|_{(\ker T)^\perp} v = 0$. Hence $Tv = 0$ and thus $v \in (\ker T) \cap (\ker T)^\perp$, which implies that $v = 0$ [by 6.52(d)]. Hence $\ker\big(T|_{(\ker T)^\perp}\big) = \{0\}$, which implies that $T|_{(\ker T)^\perp}$ is injective, as desired.

Clearly $\text{im}\big(T|_{(\ker T)^\perp}\big) \subseteq \text{im}\, T$. To prove the inclusion in the other direction, suppose $w \in \text{im}\, T$. Hence there exists $v \in V$ such that $w = Tv$. There exist $u \in \ker T$ and $x \in (\ker T)^\perp$ such that $v = u + x$ (by 6.53). Now

$$T|_{(\ker T)^\perp} x = Tx = Tv - Tu = w - 0 = w,$$

which shows that $w \in \operatorname{im} T|_{(\ker T)^\perp}$. Hence $\operatorname{im} T \subseteq \operatorname{im} T|_{(\ker T)^\perp}$, completing the proof that $\operatorname{im} T|_{(\ker T)^\perp} = \operatorname{im} T$. □

Now we can define the pseudoinverse $T^\dagger$ (pronounced "T dagger") of a linear map T. In the next definition (and from now on), think of $T|_{(\ker T)^\perp}$ as an invertible linear map from $(\ker T)^\perp$ onto $\operatorname{im} T$, as is justified by the result above.

To produce the pseudoinverse notation $T^\dagger$ in TEX, type `T^\dagger`.

Pseudoinverse, $T^\dagger$ **Definition 6.72**

Suppose that V is finite-dimensional and $T \in \operatorname{Hom}(V, W)$. The *pseudoinverse* $T^\dagger \in \operatorname{Hom}(W, V)$ of T is the linear map from W to V defined by

$$T^\dagger w = \big(T|_{(\ker T)^\perp}\big)^{-1} P_{\operatorname{im} T} w$$

for each $w \in W$.

Recall that $P_{\operatorname{im} T} w = 0$ if $w \in (\operatorname{im} T)^\perp$ and $P_{\operatorname{im} T} w = w$ if $w \in \operatorname{im} T$. Thus if $w \in (\operatorname{im} T)^\perp$, then $T^\dagger w = 0$, and if $w \in \operatorname{im} T$, then $T^\dagger w$ is the unique element of $(\ker T)^\perp$ such that $T(T^\dagger w) = w$.

The pseudoinverse behaves much like an inverse, as we will see.

Algebraic properties of the pseudoinverse **Result 6.73**

Suppose V is finite-dimensional and $T \in \operatorname{Hom}(V, W)$.

(a) If T is invertible, then $T^\dagger = T^{-1}$.

(b) $T T^\dagger = P_{\operatorname{im} T}$ = the orthogonal projection of W onto $\operatorname{im} T$.

(c) $T^\dagger T = P_{(\ker T)^\perp}$ = the orthogonal projection of V onto $(\ker T)^\perp$.

Proof.

(a) Suppose T is invertible. Then $(\ker T)^\perp = V$ and $\operatorname{im} T = W$. Thus $T|_{(\ker T)^\perp} = T$ and $P_{\operatorname{im} T}$ is the identity operator on W. Hence $T^\dagger = T^{-1}$.

(b) Suppose $w \in \operatorname{im} T$. Thus

$$T T^\dagger w = T\big(T|_{(\ker T)^\perp}\big)^{-1} w = w = P_{\operatorname{im} T} w.$$

If $w \in (\operatorname{im} T)^\perp$, then $T^\dagger w = 0$. Hence $T T^\dagger w = 0 = P_{\operatorname{im} T} w$. Thus $T T^\dagger$ and $P_{\operatorname{im} T}$ agree on $\operatorname{im} T$ and on $(\operatorname{im} T)^\perp$. Hence these two linear maps are equal (by 6.53).

(c) Suppose $v \in (\ker T)^\perp$. Because $Tv \in \operatorname{im} T$, the definition of $T^\dagger$ shows that

$$T^\dagger(Tv) = \big(T|_{(\ker T)^\perp}\big)^{-1}(Tv) = v = P_{(\ker T)^\perp} v.$$

If $v \in \ker T$, then $T^\dagger T v = 0 = P_{(\ker T)^\perp} v$. Thus $T^\dagger T$ and $P_{(\ker T)^\perp}$ agree on $(\ker T)^\perp$ and on $\ker T$. Hence these two linear maps are equal (by 6.53). □

Suppose that $T \in \operatorname{Hom}(V, W)$. If T is surjective, then $T T^\dagger$ is the identity operator on W, as follows from (b) in the result above. If T is injective, then $T^\dagger T$ is the identity

The pseudoinverse is also called the Moore–Penrose inverse.

operator on V, as follows from (c) in the result above. For additional algebraic properties of the pseudoinverse, see Exercises 19–23.

For $T \in \text{Hom}(V, W)$ and $w \in W$, we now return to the problem of finding $v \in V$ that solves the equation

$$Tv = w.$$

As we noted earlier, if T is invertible, then $v = T^{-1}w$ is the unique solution, but if T is not invertible, then T^{-1} is not defined. However, the pseudoinverse $T^\dagger$ is defined. Taking $v = T^\dagger w$ makes Tv as close to w as possible, as shown by (a) of the next result. Thus the pseudoinverse provides what is called a *best fit* to the equation above.

Among all vectors $v \in V$ that make Tv as close as possible to w, the vector $T^\dagger w$ has the smallest norm, as shown by combining (b) in the next result with the condition for equality in (a).

Result 6.74 **Pseudoinverse provides best approximate solution or best solution**

Suppose V is finite-dimensional, $T \in \text{Hom}(V, W)$, and $w \in W$.

(a) If $v \in V$, then

$$\big\|T(T^\dagger w) - w\big\| \le \|Tv - w\|,$$

with equality if and only if $v \in T^\dagger w + \ker T$.

(b) If $v \in T^\dagger w + \ker T$, then

$$\big\|T^\dagger w\big\| \le \|v\|,$$

with equality if and only if $v = T^\dagger w$.

Proof.

(a) Suppose $v \in V$. Then

$$Tv - w = (Tv - TT^\dagger w) + (TT^\dagger w - w).$$

The first term in parentheses above is in $\text{im}\, T$. Because the operator $TT^\dagger$ is the orthogonal projection of W onto $\text{im}\, T$ [by 6.73(b)], the second term in parentheses above is in $(\text{im}\, T)^\perp$ [see 6.61(f)].

Thus the Pythagorean theorem implies the desired inequality that the norm of the second term in parentheses above is less than or equal to $\|Tv - w\|$, with equality if and only if the first term in parentheses above equals 0. Hence we have equality if and only if $v - T^\dagger w \in \ker T$, which is equivalent to the statement that $v \in T^\dagger w + \ker T$, completing the proof of (a).

(b) Suppose $v \in T^\dagger w + \ker T$. Hence $v - T^\dagger w \in \ker T$. Now

$$v = (v - T^\dagger w) + T^\dagger w.$$

The definition of $T^\dagger$ implies that $T^\dagger w \in (\ker T)^\perp$. Thus the Pythagorean theorem implies that $\big\|T^\dagger w\big\| \le \|v\|$, with equality if and only if $v = T^\dagger w$. □

A formula for $T^\dagger$ will be given in the next chapter (see 7.78).

Example 6.75 **Pseudoinverse of a linear map from $\mathbb{F}^4$ to $\mathbb{F}^3$**

Suppose $T \in \text{Hom}(\mathbb{F}^4, \mathbb{F}^3)$ is defined by

$$T(a, b, c, d) = (a + b + c,\ 2c + d,\ 0).$$

This linear map is neither injective nor surjective, but we can compute its pseudoinverse. To do this, first note that $\text{im}\, T = \{(x, y, 0) : x, y \in \mathbb{F}\}$. Thus

$$P_{\text{im}\, T}(x, y, z) = (x, y, 0)$$

for each $(x, y, z) \in \mathbb{F}^3$. Also,

$$\ker T = \{(a, b, c, d) \in \mathbb{F}^4 : a + b + c = 0 \text{ and } 2c + d = 0\}.$$

The list $(-1, 1, 0, 0)$, $(-1, 0, 1, -2)$ of two vectors in $\ker T$ spans $\ker T$ because if $(a, b, c, d) \in \ker T$ then

$$(a, b, c, d) = b(-1, 1, 0, 0) + c(-1, 0, 1, -2).$$

Because the list $(-1, 1, 0, 0)$, $(-1, 0, 1, -2)$ is linearly independent, this list is a basis of $\ker T$.

Now suppose $(x, y, z) \in \mathbb{F}^3$. Then

$$T^\dagger(x, y, z) = \big(T|_{(\ker T)^\perp}\big)^{-1} P_{\text{im}\, T}(x, y, z) = \big(T|_{(\ker T)^\perp}\big)^{-1}(x, y, 0). \qquad 6.76$$

The right side of the equation above is the vector $(a, b, c, d) \in \mathbb{F}^4$ such that $T(a, b, c, d) = (x, y, 0)$ and $(a, b, c, d) \in (\ker T)^\perp$. In other words, a, b, c, d must satisfy the following equations:

$$\begin{aligned} a + b + c &= x \\ 2c + d &= y \\ -a + b &= 0 \\ -a + c - 2d &= 0, \end{aligned}$$

where the first two equations are equivalent to the equation $T(a, b, c, d) = (x, y, 0)$ and the last two equations come from the condition for (a, b, c, d) to be orthogonal to each of the basis vectors $(-1, 1, 0, 0)$, $(-1, 0, 1, -2)$ in this basis of $\ker T$. Thinking of x and y as constants and a, b, c, d as unknowns, we can solve the system above of four equations in four unknowns, getting

$$a = (1/11)(5x-2y),\ b = (1/11)(5x-2y),\ c = (1/11)(x+4y),\ d = (1/11)(-2x+3y).$$

Hence 6.76 tells us that

$$T^\dagger(x, y, z) = (1/11)(5x - 2y,\ 5x - 2y,\ x + 4y,\ -2x + 3y).$$

The formula above for $T^\dagger$ shows that $TT^\dagger(x, y, z) = (x, y, 0)$ for all $(x, y, z) \in \mathbb{F}^3$, which illustrates the equation $TT^\dagger = P_{\text{im}\, T}$ from 6.73(b).

Exercises

6C.1 Suppose $v_1, \dots, v_m \in V$. Prove that

$$\{v_1, \dots, v_m\}^\perp = \big(\operatorname{span}(v_1, \dots, v_m)\big)^\perp.$$

6C.2 Suppose U is a subspace of V with basis $u_1, \dots, u_m$ and

$$u_1, \dots, u_m, v_1, \dots, v_n$$

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 $U^\perp$.

6C.3 Suppose U is the subspace of $\mathbb{R}^4$ 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 $U^\perp$.

6C.4 Suppose $e_1, \dots, e_n$ is a list of vectors in V with $\|e_k\| = 1$ for each $k = 1, \dots, n$ and

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2$$

for all $v \in V$. Prove that $e_1, \dots, e_n$ is an orthonormal basis of V.

This exercise provides a converse to 6.34(b).

6C.5 Suppose that 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.

6C.6 Suppose V is finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Show that

$$T = T P_{(\ker T)^\perp} = P_{\operatorname{im} T} T.$$

6C.7 Suppose that X and Y are finite-dimensional subspaces of V. Prove that $P_X P_Y = 0$ if and only if $\langle x, y\rangle = 0$ for all $x \in X$ and all $y \in Y$.

6C.8 Suppose U is a finite-dimensional subspace of V and $v \in V$. Define a linear functional $\varphi: U \to \mathbb{F}$ by

$$\varphi(u) = \langle u, v \rangle$$

for all $u \in U$. By the Riesz representation theorem (6.46) as applied to the inner product space U, there exists a unique vector $w \in U$ such that

$$\varphi(u) = \langle u, w \rangle$$

for all $u \in U$. Show that $w = P_U v$.

6C.9 Suppose V is finite-dimensional. Suppose $P \in \text{End}(V)$ is such that $P^2 = P$ and every vector in $\ker P$ is orthogonal to every vector in $\text{im } P$. Prove that there exists a subspace U of V such that $P = P_U$.

6C.10 Suppose V is finite-dimensional and $P \in \text{End}(V)$ 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 $P = P_U$.

6C.11 Suppose $T \in \text{End}(V)$ and U is a finite-dimensional subspace of V. Prove that

$$U \text{ is invariant under } T \iff P_U T P_U = T P_U.$$

6C.12 Suppose V is finite-dimensional, $T \in \text{End}(V)$, and U is a subspace of V. Prove that

$$U \text{ and } U^\perp \text{ are both invariant under } T \iff P_U T = T P_U.$$

6C.13 Suppose $\mathbb{F} = \mathbb{R}$ and V is finite-dimensional. For each $v \in V$, let φ_v denote the linear functional on V defined by

$$\varphi_v(u) = \langle u, v \rangle$$

for all $u \in V$.

(a) Show that $v \mapsto \varphi_v$ is an injective linear map from V to V'.

(b) Use (a) and a dimension-counting argument to show that $v \mapsto \varphi_v$ is an isomorphism from V onto V'.

The purpose of this exercise is to give an alternative proof of the Riesz representation theorem (6.46 and 6.62) when $\mathbb{F} = \mathbb{R}$. Thus you should not use the Riesz representation theorem as a tool in your solution.

6C.14 Suppose $e_1, \dots, e_n$ is an orthonormal basis of V. Explain why the dual basis (see 3.112) of $e_1, \dots, e_n$ is $e_1, \dots, e_n$ under the identification of V' with V provided by the Riesz representation theorem (6.62).

6C.15 In $\mathbb{R}^4$, let

$$U = \operatorname{span}\big((1, 1, 0, 0), (1, 1, 1, 2)\big).$$

Find $u \in U$ such that $\|u - (1, 2, 3, 4)\|$ is as small as possible.

6C.16 Suppose $C[-1, 1]$ is the vector space of continuous real-valued functions on the interval $[-1, 1]$ with inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} fg$$

for all $f, g \in C[-1, 1]$. Let U be the subspace of $C[-1, 1]$ defined by

$$U = \{f \in C[-1, 1] : f(0) = 0\}.$$

(a) Show that $U^\perp = \{0\}$.

(b) Show that 6.53 and 6.56 do not hold without the finite-dimensional hypothesis.

6C.17 Find $p \in \mathbb{R}[X]_{\le 3}$ such that $p(0) = 0$, $p'(0) = 0$, and $\int_0^1 |2 + 3x - p(x)|^2 \, dx$ is as small as possible.

6C.18 Find $p \in \mathbb{R}[X]_{\le 5}$ that makes $\int_{-\pi}^{\pi} |\sin x - p(x)|^2 \, dx$ as small as possible.

The polynomial 6.69 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 should help.

6C.19 Suppose V is finite-dimensional and $P \in \operatorname{End}(V)$ is an orthogonal projection of V onto some subspace of V. Prove that $P^\dagger = P$.

6C.20 Suppose V is finite-dimensional and $T \in \operatorname{Hom}(V, W)$. Show that

$$\ker T^\dagger = (\operatorname{im} T)^\perp \quad \text{and} \quad \operatorname{im} T^\dagger = (\ker T)^\perp.$$

6C.21 Suppose $T \in \operatorname{Hom}(\mathbb{F}^3, \mathbb{F}^2)$ is defined by

$$T(a, b, c) = (a + b + c, 2b + 3c).$$

(a) For $(x, y) \in \mathbb{F}^2$, find a formula for $T^\dagger(x, y)$.

(b) Verify that the equation $TT^\dagger = P_{\operatorname{im} T}$ from 6.73(b) holds with the formula for $T^\dagger$ obtained in (a).

(c) Verify that the equation $T^\dagger T = P_{(\ker T)^\perp}$ from 6.73(c) holds with the formula for $T^\dagger$ obtained in (a).

6C.22 Suppose V is finite-dimensional and $T \in \text{Hom}(V, W)$. Prove that

$$TT^\dagger T = T \quad \text{and} \quad T^\dagger T T^\dagger = T^\dagger.$$

Both formulas above clearly hold if T is invertible because in that case we can replace $T^\dagger$ with T^{-1}.

6C.23 Suppose V and W are finite-dimensional and $T \in \text{Hom}(V, W)$. Prove that

$$(T^\dagger)^\dagger = T.$$

The equation above is analogous to the equation $(T^{-1})^{-1} = T$ that holds if T is invertible.

Operators on Inner Product Spaces

7

The deepest results related to inner product spaces deal with the subject to which we now turn—linear maps and operators on inner product spaces. As we will see, good theorems can be proved by exploiting properties of the adjoint.

The hugely important spectral theorem will provide a complete description of self-adjoint operators on real inner product spaces and of normal operators on complex inner product spaces. We will then use the spectral theorem to help understand positive operators and unitary operators, which will lead to unitary matrices and matrix factorizations. The spectral theorem will also lead to the popular singular value decomposition, which will lead to the polar decomposition.

The most important results in the rest of this book are valid only in finite dimensions. Thus from now on we assume that V and W are finite-dimensional.

Standing assumptions for this chapter — Definition

- $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$.
- V and W are nonzero finite-dimensional inner product spaces over $\mathbb{F}$.

7A Self-Adjoint and Normal Operators

Adjoints

Adjoint, T^* — Definition 7.1

Suppose $T \in \text{Hom}(V, W)$. 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. In case you encounter the second meaning elsewhere, be warned that the two meanings for adjoint are unrelated to each other.

To see why the definition above makes sense, suppose $T \in \text{Hom}(V, W)$. Fix $w \in W$. Consider the linear functional

$$v \mapsto \langle Tv, w\rangle$$

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 (6.46), 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 $v \in V$.

In the equation above, the inner product on the left takes place in W and the inner product on the right takes place in V. However, we use the same notation $\langle \cdot, \cdot \rangle$ for both inner products.

Example 7.2 **Adjoint of a linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$**

Define $T: \mathbb{R}^3 \to \mathbb{R}^2$ by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3,\ 2x_1).$$

To compute T^*, suppose $(x_1, x_2, x_3) \in \mathbb{R}^3$ and $(y_1, y_2) \in \mathbb{R}^2$. Then

$$\begin{aligned}\big\langle T(x_1, x_2, x_3), (y_1, y_2)\big\rangle &= \big\langle (x_2 + 3x_3,\ 2x_1), (y_1, y_2)\big\rangle \\ &= x_2y_1 + 3x_3y_1 + 2x_1y_2 \\ &= \big\langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\big\rangle.\end{aligned}$$

The equation above and the definition of the adjoint imply that

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

Example 7.3 **Adjoint of a linear map with range of dimension at most 1**

Fix $u \in V$ and $x \in W$. Define $T \in \text{Hom}(V, W)$ by

$$Tv = \langle v, u\rangle x$$

for each $v \in V$. To compute T^*, suppose $v \in V$ and $w \in W$. Then

$$\begin{aligned}\langle Tv, w\rangle &= \big\langle \langle v, u\rangle x, w\big\rangle \\ &= \langle v, u\rangle\langle x, w\rangle \\ &= \big\langle v, \langle w, x\rangle u\big\rangle.\end{aligned}$$

Thus

$$T^*w = \langle w, x\rangle u.$$

The two examples above and the proof below use a common technique for computing T^*: start with a formula for $\langle Tv, w\rangle$ then manipulate it to get just v in the first slot; the entry in the second slot will then be T^*w.

In the two examples above, T^* turned out to be not just a function from W to V but a linear map from W to V. This behavior is true in general, as shown by the next result.

Adjoint of a linear map is a linear map **Result 7.4**

If $T \in \text{Hom}(V, W)$, then $T^* \in \text{Hom}(W, V)$.

Proof. Suppose $T \in \text{Hom}(V, W)$. If $v \in V$ and $w_1, w_2 \in W$, then

$$\begin{aligned}\langle Tv, w_1 + w_2\rangle &= \langle Tv, w_1\rangle + \langle Tv, w_2\rangle \\ &= \langle v, T^*w_1\rangle + \langle v, T^*w_2\rangle \\ &= \langle v, T^*w_1 + T^*w_2\rangle.\end{aligned}$$

The equation above shows that

$$T^*(w_1 + w_2) = T^*w_1 + T^*w_2.$$

If $v \in V$, $\lambda \in \mathbb{F}$, and $w \in W$, then

$$\begin{aligned}\langle Tv, \lambda w\rangle &= \overline{\lambda}\langle Tv, w\rangle \\ &= \overline{\lambda}\langle v, T^*w\rangle \\ &= \langle v, \lambda T^*w\rangle.\end{aligned}$$

The equation above shows that

$$T^*(\lambda w) = \lambda T^*w.$$

Thus T^* is a linear map, as desired. □

Properties of the adjoint **Result 7.5**

Suppose $T \in \text{Hom}(V, W)$. Then

(a) $(S + T)^* = S^* + T^*$ for all $S \in \text{Hom}(V, W)$;

(b) $(\lambda T)^* = \overline{\lambda}T^*$ for all $\lambda \in \mathbb{F}$;

(c) $(T^*)^* = T$;

(d) $(ST)^* = T^*S^*$ for all $S \in \text{Hom}(W, U)$ (here U is a finite-dimensional inner product space over $\mathbb{F}$);

(e) $I^* = I$, where I is the identity operator on V;

(f) if T is invertible, then T^* is invertible and $(T^*)^{-1} = (T^{-1})^*$.

Proof. Suppose $v \in V$ and $w \in W$.

(a) If $S \in \text{Hom}(V, W)$, then

$$\begin{aligned}\big\langle (S + T)v, w\big\rangle &= \langle Sv, w\rangle + \langle Tv, w\rangle \\ &= \langle v, S^*w\rangle + \langle v, T^*w\rangle \\ &= \langle v, S^*w + T^*w\rangle.\end{aligned}$$

Thus $(S + T)^*w = S^*w + T^*w$, as desired.

(b) If $\lambda \in \mathbb{F}$, then

$$\big\langle (\lambda T)v, w\big\rangle = \lambda\langle Tv, w\rangle = \lambda\langle v, T^*w\rangle = \langle v, \overline{\lambda}T^*w\rangle.$$

Thus $(\lambda T)^*w = \overline{\lambda}T^*w$, as desired.

(c) We have

$$\langle T^*w, v\rangle = \overline{\langle v, T^*w\rangle} = \overline{\langle Tv, w\rangle} = \langle w, Tv\rangle.$$

Thus $(T^*)^*v = Tv$, as desired.

(d) Suppose $S \in \mathrm{Hom}(W, U)$ and $u \in U$. Then

$$\big\langle (ST)v, u\big\rangle = \big\langle S(Tv), u\big\rangle = \langle Tv, S^*u\rangle = \big\langle v, T^*(S^*u)\big\rangle.$$

Thus $(ST)^*u = T^*(S^*u)$, as desired.

(e) Suppose $u \in V$. Then

$$\langle Iu, v\rangle = \langle u, v\rangle.$$

Thus $I^*v = v$, as desired.

(f) Suppose T is invertible. Take adjoints of both sides of the equation $T^{-1}T = I$, then use (d) and (e) to show that $T^*(T^{-1})^* = I$. Similarly, the equation $TT^{-1} = I$ implies $(T^{-1})^*T^* = I$. Thus $(T^{-1})^*$ is the inverse of T^*, as desired. □

If $\mathbb{F} = \mathbb{R}$, then the map $T \mapsto T^*$ is a linear map from $\mathrm{Hom}(V, W)$ to $\mathrm{Hom}(W, V)$, as follows from (a) and (b) of the result above. However, if $\mathbb{F} = \mathbb{C}$, then this map is not linear because of the complex conjugate that appears in (b).

The next result shows the relationship between the null space and the range of a linear map and its adjoint.

Result 7.6

Null space and range of T^*

Suppose $T \in \mathrm{Hom}(V, W)$. Then

(a) $\ker T^* = (\mathrm{im}\, T)^\perp$;

(b) $\mathrm{im}\, T^* = (\ker T)^\perp$;

(c) $\ker T = (\mathrm{im}\, T^*)^\perp$;

(d) $\mathrm{im}\, T = (\ker T^*)^\perp$.

Proof. We begin by proving (a). Let $w \in W$. Then

$$\begin{aligned} w \in \ker T^* &\iff T^*w = 0 \\ &\iff \langle v, T^*w\rangle = 0 \text{ for all } v \in V \\ &\iff \langle Tv, w\rangle = 0 \text{ for all } v \in V \\ &\iff w \in (\mathrm{im}\, T)^\perp. \end{aligned}$$

Thus $\ker T^* = (\mathrm{im}\, T)^\perp$, proving (a).

If we take the orthogonal complement of both sides of (a), we get (d), where we have used 6.56. Replacing T with T^* in (a) gives (c), where we have used 7.5(c). Finally, replacing T with T^* in (d) gives (b). □

As we will soon see, the next definition is intimately connected to the matrix of the adjoint of a linear map.

Conjugate transpose, A^* — Definition 7.7

The *conjugate transpose* of an m-by-n matrix A is the n-by-m matrix A^* obtained by interchanging the rows and columns and then taking the complex conjugate of each entry. In other words, if $j \in \{1, \dots, n\}$ and $k \in \{1, \dots, m\}$, then

$$(A^*)_{j,k} = \overline{A_{k,j}}.$$

If a matrix A has only real entries, then $A^* = A^T$, where A^T denotes the transpose of A (the matrix obtained by interchanging the rows and the columns).

The adjoint of a linear map does not depend on a choice of basis. Thus we frequently emphasize adjoints of linear maps instead of transposes or conjugate transposes of matrices.

Example 7.8 **Conjugate transpose of a 2-by-3 matrix**

The conjugate transpose of the 2-by-3 matrix $\begin{pmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{pmatrix}$ is the 3-by-2 matrix

$$\begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}.$$

The next result shows how to compute the matrix of T^* from the matrix of T. **Caution:** With respect to nonorthonormal bases, the matrix of T^* does not necessarily equal the conjugate transpose of the matrix of T.

Matrix of T^* equals conjugate transpose of matrix of T — Result 7.9

Let $T \in \text{Hom}(V, W)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $f_1, \dots, f_m$ is an orthonormal basis of W. Then $\text{Mat}\big(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n)\big)$ is the conjugate transpose of $\text{Mat}\big(T, (e_1, \dots, e_n), (f_1, \dots, f_m)\big)$. In other words,

$$\text{Mat}(T^*) = \big(\text{Mat}(T)\big)^*.$$

Proof. In this proof, we will write $\text{Mat}(T)$ and $\text{Mat}(T^*)$ instead of the longer expressions $\text{Mat}\big(T, (e_1, \dots, e_n), (f_1, \dots, f_m)\big)$ and $\text{Mat}\big(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n)\big)$.

Recall that we obtain the kth column of $\text{Mat}(T)$ by writing Te_k as a linear combination of the f_j's; the scalars used in this linear combination then become the kth column of $\text{Mat}(T)$. Because $f_1, \dots, f_m$ is an orthonormal basis of W, we know how to write Te_k as a linear combination of the f_j's [see 6.34(a)]:

$$Te_k = \langle Te_k, f_1 \rangle f_1 + \cdots + \langle Te_k, f_m \rangle f_m.$$

Thus

$$\text{the entry in row } j, \text{ column } k, \text{ of } \text{Mat}(T) \text{ is } \langle Te_k, f_j \rangle.$$

In the statement above, replace T with T^* and interchange $e_1, \dots, e_n$ and $f_1, \dots, f_m$. This shows that the entry in row j, column k, of $\text{Mat}(T^*)$ 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 $\text{Mat}(T)$. Thus $\text{Mat}(T^*) = \big(\text{Mat}(T)\big)^*$. □

The Riesz representation theorem as stated in 6.62 provides an identification of V with its dual space V' defined in 3.110. Under this identification, the orthogonal complement $U^\perp$ of a subset $U \subseteq V$ corresponds to the annihilator U^0 of U. If U is a subspace of V, then the formulas for the dimensions of $U^\perp$ and U^0 become identical under this identification—see 3.125 and 6.55.

Suppose $T: V \to W$ is a linear map. Under the identification of V with V' and the identification of W with W', the adjoint map $T^*: W \to V$ corresponds to the dual map $T': W' \to V'$ defined in 3.118, as Exercise 32 asks you to verify. Under this identification, the formulas for $\ker T^*$ and $\operatorname{im} T^*$ [7.6(a) and (b)] then become identical to the formulas for $\ker T'$ and $\operatorname{im} T'$ [3.128(a) and 3.130(b)]. Furthermore, the theorem about the matrix of T^* (7.9) is analogous to the theorem about the matrix of T' (3.132).

Because orthogonal complements and adjoints are easier to deal with than annihilators and dual maps, there is no need to work with annihilators and dual maps in the context of inner product spaces.

Self-Adjoint Operators

Now we switch our attention to operators on inner product spaces. Instead of considering linear maps from V to W, we will focus on linear maps from V to V; recall that such linear maps are called operators.

Self-adjoint Definition 7.10

An operator $T \in \operatorname{End}(V)$ is called *self-adjoint* if $T = T^*$.

If $T \in \operatorname{End}(V)$ and $e_1, \dots, e_n$ is an orthonormal basis of V, then T is self-adjoint if and only if $\operatorname{Mat}\big(T, (e_1, \dots, e_n)\big) = \operatorname{Mat}\big(T, (e_1, \dots, e_n)\big)^*$, as follows from 7.9.

Example 7.11 **Determining whether T is self-adjoint from its matrix**

Suppose $c \in \mathbb{F}$ and T is the operator on $\mathbb{F}^2$ whose matrix (with respect to the standard basis) is

$$\operatorname{Mat}(T) = \begin{pmatrix} 2 & c \\ 3 & 7 \end{pmatrix}.$$

The matrix of T^* (with respect to the standard basis) is

$$\operatorname{Mat}(T^*) = \begin{pmatrix} 2 & 3 \\ \overline{c} & 7 \end{pmatrix}.$$

Thus $\operatorname{Mat}(T) = \operatorname{Mat}(T^*)$ if and only if $c = 3$. Hence the operator T is self-adjoint if and only if $c = 3$.

A good analogy to keep in mind is that the adjoint on $\operatorname{End}(V)$ plays a role similar to that of the complex conjugate on $\mathbb{C}$. A complex number z is real if and only if $z = \overline{z}$; thus a self-adjoint operator ($T = T^*$) is analogous to a real number.

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.

An operator $T \in \operatorname{End}(V)$ is self-adjoint if and only if

$$\langle Tv, w \rangle = \langle v, Tw \rangle$$

for all $v, w \in V$.

If $\mathbb{F} = \mathbb{R}$, then by definition every eigenvalue is real, so the next result is interesting only when $\mathbb{F} = \mathbb{C}$.

Eigenvalues of self-adjoint operators — **Result 7.12**

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

$$\lambda\|v\|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\|v\|^2.$$

Thus $\lambda = \overline{\lambda}$, which means that λ is real, as desired. □

The next result is false for real inner product spaces. As an example, consider the operator $T \in \text{End}(\mathbb{R}^2)$ that is a counterclockwise rotation of 90° around the origin; thus $T(x, y) = (-y, x)$. Notice that Tv is orthogonal to v for every $v \in \mathbb{R}^2$, even though $T \neq 0$.

Tv is orthogonal to v for all v $\iff$ $T = 0$ (assuming $\mathbb{F} = \mathbb{C}$) — **Result 7.13**

Suppose V is a complex inner product space and $T \in \text{End}(V)$. Then

$$\langle Tv, v\rangle = 0 \text{ for every } v \in V \iff T = 0.$$

Proof. If $u, w \in V$, then

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} + \frac{\langle T(u+iw), u+iw\rangle - \langle T(u-iw), u-iw\rangle}{4}\, i,$$

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 $v \in V$.

Now suppose $\langle Tv, v\rangle = 0$ for every $v \in V$. Then the equation above implies that $\langle Tu, w\rangle = 0$ for all $u, w \in V$, which then implies that $Tu = 0$ for every $u \in V$ (take $w = Tu$). Hence $T = 0$, as desired. □

The next result provides another good 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.

$\langle Tv, v\rangle$ is real for all v $\iff$ T is self-adjoint (assuming $\mathbb{F} = \mathbb{C}$) — **Result 7.14**

Suppose V is a complex inner product space and $T \in \text{End}(V)$. Then

$$T \text{ is self-adjoint} \iff \langle Tv, v\rangle \in \mathbb{R} \text{ for every } v \in V.$$

Proof. If $v \in V$, then

$$\langle T^*v, v\rangle = \overline{\langle v, T^*v\rangle} = \overline{\langle Tv, v\rangle}. \tag{7.15}$$

Now

$$\begin{aligned} T \text{ is self-adjoint} &\iff T - T^* = 0 \\ &\iff \big\langle (T - T^*)v, v \big\rangle = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle \in \mathbb{R} \text{ for every } v \in V, \end{aligned}$$

where the second equivalence follows from 7.13 as applied to $T - T^*$ and the third equivalence follows from 7.15. □

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.

Result 7.16 — **T self-adjoint and $\langle Tv, v\rangle = 0$ for all v $\iff$ $T = 0$**

Suppose T is a self-adjoint operator on V. Then

$$\langle Tv, v\rangle = 0 \text{ for every } v \in V \iff T = 0.$$

Proof. We have already proved this (without the hypothesis that T is self-adjoint) when V is a complex inner product space (see 7.13). Thus we can assume that V is a real inner product space. If $u, w \in V$, then

$$\langle Tu, w\rangle = \frac{\big\langle T(u+w), u+w\big\rangle - \big\langle T(u-w), u-w\big\rangle}{4}, \tag{7.17}$$

as can be 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.

Now suppose $\langle Tv, v\rangle = 0$ for every $v \in V$. Because each term on the right side of 7.17 is of the form $\langle Tv, v\rangle$ for appropriate v, this implies that $\langle Tu, w\rangle = 0$ for all $u, w \in V$. This implies that $Tu = 0$ for every $u \in V$ (take $w = Tu$). Hence $T = 0$, as desired. □

Normal Operators

Definition 7.18 — **Normal**

- An operator on an inner product space is called *normal* if it commutes with its adjoint.
- In other words, $T \in \text{End}(V)$ is normal if $TT^* = T^*T$.

Every self-adjoint operator is normal, because if T is self-adjoint then $T^* = T$ and hence T commutes with T^*.

Example 7.19 **An operator that is normal but not self-adjoint**

Let T be the operator on $\mathbb{F}^2$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

Thus $T(w, z) = (2w - 3z, 3w + 2z)$.

This operator T 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

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}, \quad \text{which equals} \quad \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Similarly, the matrix of T^*T equals

$$\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}, \quad \text{which equals} \quad \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Because TT^* and T^*T have the same matrix, we see that $TT^* = T^*T$. Thus T is normal.

In the next section we will see why normal operators are worthy of special attention. The next result provides a useful characterization of normal operators.

Result 7.20 **T is normal if and only if Tv and T^*v have the same norm**

Suppose $T \in \text{End}(V)$. Then

$$T \text{ is normal} \iff \|Tv\| = \|T^*v\| \text{ for every } v \in V.$$

Proof. We have

$$\begin{aligned} T \text{ is normal} &\iff T^*T - TT^* = 0 \\ &\iff \big\langle (T^*T - TT^*)v, v \big\rangle = 0 \text{ for every } v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \text{ for every } v \in V \\ &\iff \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle \text{ for every } v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \text{ for every } v \in V \\ &\iff \|Tv\| = \|T^*v\| \text{ for every } v \in V, \end{aligned}$$

where we used 7.16 to establish the second equivalence (note that the operator $T^*T - TT^*$ is self-adjoint). □

The next result presents several consequences of the result above. Compare (e) of the next result to Exercise 3. 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, (e) of the next result implies that a normal operator and its adjoint have the same eigenvectors.

Range, null space, and eigenvectors of a normal operator **Result 7.21**

Suppose $T \in \text{End}(V)$ is normal. Then

(a) $\ker T = \ker T^*$;

(b) $\text{im}\, T = \text{im}\, T^*$;

(c) $V = \ker T \oplus \text{im}\, T$;

(d) $T - \lambda I$ is normal for every $\lambda \in \mathbb{F}$;

(e) if $v \in V$ and $\lambda \in \mathbb{F}$, then $Tv = \lambda v$ if and only if $T^*v = \overline{\lambda} v$.

Proof.

(a) Suppose $v \in V$. Then

$$v \in \ker T \iff \|Tv\| = 0 \iff \|T^*v\| = 0 \iff v \in \ker T^*,$$

where the middle equivalence above follows from 7.20. Thus $\ker T = \ker T^*$.

(b) We have

$$\text{im}\, T = (\ker T^*)^\perp = (\ker T)^\perp = \text{im}\, T^*,$$

where the first equality comes from 7.6(d), the second equality comes from (a) in this result, and the third equality comes from 7.6(b).

(c) We have

$$V = (\ker T) \oplus (\ker T)^\perp = \ker T \oplus \text{im}\, T^* = \ker T \oplus \text{im}\, T,$$

where the first equality comes from 6.53, the second equality comes from 7.6(b), and the third equality comes from (b) in this result.

(d) Suppose $\lambda \in \mathbb{F}$. Then

$$\begin{aligned}(T - \lambda I)(T - \lambda I)^* &= (T - \lambda I)(T^* - \overline{\lambda} I)\\ &= TT^* - \overline{\lambda} T - \lambda T^* + |\lambda|^2 I\\ &= T^*T - \overline{\lambda} T - \lambda T^* + |\lambda|^2 I\\ &= (T^* - \overline{\lambda} I)(T - \lambda I)\\ &= (T - \lambda I)^*(T - \lambda I).\end{aligned}$$

Thus $T - \lambda I$ commutes with its adjoint. Hence $T - \lambda I$ is normal.

(e) Suppose $v \in V$ and $\lambda \in \mathbb{F}$. Then (d) and 7.20 imply that

$$\|(T - \lambda I)v\| = \|(T - \lambda I)^* v\| = \|(T^* - \overline{\lambda} I)v\|.$$

Thus $\|(T - \lambda I)v\| = 0$ if and only if $\|(T^* - \overline{\lambda} I)v\| = 0$. Hence $Tv = \lambda v$ if and only if $T^*v = \overline{\lambda} v$. □

Because every self-adjoint operator is normal, the next result applies in particular to self-adjoint operators.

Orthogonal eigenvectors for normal operators — Result 7.22

Suppose $T \in \text{End}(V)$ 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(e) we have $T^*v = \overline{\beta} v$. Thus

$$\begin{aligned}(\alpha - \beta)\langle u, v\rangle &= \langle \alpha u, v\rangle - \langle u, \overline{\beta} v\rangle \\ &= \langle Tu, v\rangle - \langle u, T^*v\rangle \\ &= 0.\end{aligned}$$

Because $\alpha \neq \beta$, the equation above implies that $\langle u, v\rangle = 0$. Thus u and v are orthogonal, as desired. □

As stated here, the next result makes sense only when $\mathbb{F} = \mathbb{C}$. However, see Exercise 12 for a version that makes sense when $\mathbb{F} = \mathbb{C}$ and when $\mathbb{F} = \mathbb{R}$.

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Under the analogy between $\text{End}(V)$ and $\mathbb{C}$, with the adjoint on $\text{End}(V)$ playing a similar role to that of the complex conjugate on $\mathbb{C}$, the operators A and B as defined by 7.24 correspond to the real and imaginary parts of T. Thus the informal title of the result below should make sense.

T is normal $\iff$ the real and imaginary parts of T commute — Result 7.23

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Then T is normal if and only if there exist commuting self-adjoint operators A and B such that $T = A + iB$.

Proof. First suppose T is normal. Let

$$A = \frac{T + T^*}{2} \quad \text{and} \quad B = \frac{T - T^*}{2i}. \tag{7.24}$$

Then A and B are self-adjoint and $T = A + iB$. A quick computation shows that

$$AB - BA = \frac{T^*T - TT^*}{2i}. \tag{7.25}$$

Because T is normal, the right side of the equation above equals 0. Thus the operators A and B commute, as desired.

To prove the implication in the other direction, now suppose there exist commuting self-adjoint operators A and B such that $T = A + iB$. Then $T^* = A - iB$. Adding the last two equations and then dividing by 2 produces the equation for A in 7.24. Subtracting the last two equations and then dividing by $2i$ produces the equation for B in 7.24. Now 7.24 implies 7.25. Because B and A commute, 7.25 implies that T is normal, as desired. □

Exercises

7A.1 Suppose n is a positive integer. Define $T \in \text{End}(\mathbb{F}^n)$ by

$$T(z_1, \dots, z_n) = (0, z_1, \dots, z_{n-1}).$$

Find a formula for $T^*(z_1, \dots, z_n)$.

7A.2 Suppose $T \in \text{Hom}(V, W)$. Prove that

$$T = 0 \iff T^* = 0 \iff T^*T = 0 \iff TT^* = 0.$$

7A.3 Suppose $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$. Prove that

$$\lambda \text{ is an eigenvalue of } T \iff \overline{\lambda} \text{ is an eigenvalue of } T^*.$$

7A.4 Suppose $T \in \text{End}(V)$ and U is a subspace of V. Prove that

$$U \text{ is invariant under } T \iff U^\perp \text{ is invariant under } T^*.$$

7A.5 Suppose $T \in \text{Hom}(V, W)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $f_1, \dots, f_m$ is an orthonormal basis of W. Prove that

$$\|Te_1\|^2 + \cdots + \|Te_n\|^2 = \|T^* f_1\|^2 + \cdots + \|T^* f_m\|^2.$$

The numbers $\|Te_1\|^2, \dots, \|Te_n\|^2$ in the equation above depend on the orthonormal basis $e_1, \dots, e_n$, but the right side of the equation does not depend on $e_1, \dots, e_n$. Thus the equation above shows that the sum on the left side does not depend on which orthonormal basis $e_1, \dots, e_n$ is used.

7A.6 Suppose $T \in \text{Hom}(V, W)$. Prove that

(a) T is injective $\iff$ T^* is surjective;

(b) T is surjective $\iff$ T^* is injective.

7A.7 Prove that if $T \in \text{Hom}(V, W)$, then

(a) $\dim \ker T^* = \dim \ker T + \dim W - \dim V$;

(b) $\dim \text{im } T^* = \dim \text{im } T$.

7A.8 Suppose A is an m-by-n matrix with entries in $\mathbb{F}$. Use (b) in Exercise 7 to prove that the row rank of A equals the column rank of A.

This exercise asks for yet another alternative proof of a result that was previously proved in 3.57 and 3.133.

7A.9 Prove that the product of two self-adjoint operators on V is self-adjoint if and only if the two operators commute.

7A.10 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that T is self-adjoint if and only if

$$\langle Tv, v\rangle = \langle T^*v, v\rangle$$

for all $v \in V$.

7A.11 Define an operator $S: \mathbb{F}^2 \to \mathbb{F}^2$ by $S(w, z) = (-z, w)$.

(a) Find a formula for S^*.

(b) Show that S is normal but not self-adjoint.

(c) Find all eigenvalues of S.

If $\mathbb{F} = \mathbb{R}$, then S is the operator on $\mathbb{R}^2$ of counterclockwise rotation by 90°.

7A.12 An operator $B \in \text{End}(V)$ is called *skew* if

$$B^* = -B.$$

Suppose that $T \in \text{End}(V)$. Prove that T is normal if and only if there exist commuting operators A and B such that A is self-adjoint, B is a skew operator, and $T = A + B$.

7A.13 Suppose $\mathbb{F} = \mathbb{R}$. Define $\mathcal{A} \in \text{End}\big(\text{End}(V)\big)$ by $\mathcal{A}T = T^*$ for all $T \in \text{End}(V)$.

(a) Find all eigenvalues of $\mathcal{A}$.

(b) Find the minimal polynomial of $\mathcal{A}$.

7A.14 Define an inner product on $\mathbb{R}[X]_{\le 2}$ by $\langle p, q\rangle = \int_0^1 pq$. Define an operator $T \in \text{End}\big(\mathbb{R}[X]_{\le 2}\big)$ by

$$T(ax^2 + bx + c) = bx.$$

(a) Show that with this inner product, the operator 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.

7A.15 Suppose $T \in \text{End}(V)$ is invertible. Prove that
(a) T is self-adjoint $\iff$ T^{-1} is self-adjoint;
(b) T is normal $\iff$ T^{-1} is normal.

7A.16 Suppose $\mathbb{F} = \mathbb{R}$.
(a) Show that the set of self-adjoint operators on V is a subspace of $\text{End}(V)$.
(b) What is the dimension of the subspace of $\text{End}(V)$ in (a) [in terms of $\dim V$]?

7A.17 Suppose $\mathbb{F} = \mathbb{C}$. Show that the set of self-adjoint operators on V is not a subspace of $\text{End}(V)$.

7A.18 Suppose $\dim V \geq 2$. Show that the set of normal operators on V is not a subspace of $\text{End}(V)$.

7A.19 Suppose $T \in \text{End}(V)$ and $\|T^*v\| \leq \|Tv\|$ for every $v \in V$. Prove that T is normal.

This exercise fails on infinite-dimensional inner product spaces, leading to what are called hyponormal operators, which have a well-developed theory.

7A.20 Suppose $P \in \text{End}(V)$ is such that $P^2 = P$. Prove that the following are equivalent.
(a) P is self-adjoint.
(b) P is normal.
(c) There is a subspace U of V such that $P = P_U$.

7A.21 Suppose $D: \mathbb{R}[X]_{\leq 8} \to \mathbb{R}[X]_{\leq 8}$ is the differentiation operator defined by $Dp = p'$. Prove that there does not exist an inner product on $\mathbb{R}[X]_{\leq 8}$ that makes D a normal operator.

7A.22 Give an example of an operator $T \in \text{End}(\mathbb{R}^3)$ such that T is normal but not self-adjoint.

7A.23 Suppose T is a normal operator on V. Suppose also that $v, w \in V$ satisfy the equations

$$\|v\| = \|w\| = 2, \qquad Tv = 3v, \qquad Tw = 4w.$$

Show that $\|T(v + w)\| = 10$.

7A.24 Suppose $T \in \text{End}(V)$ and

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

is the minimal polynomial of T. Prove that the minimal polynomial of T^* is

$$\overline{a_0} + \overline{a_1} z + \overline{a_2} z^2 + \cdots + \overline{a_{m-1}} z^{m-1} + z^m.$$

This exercise shows that the minimal polynomial of T^* equals the minimal polynomial of T if $\mathbb{F} = \mathbb{R}$.

7A.25 Suppose $T \in \text{End}(V)$. Prove that T is diagonalizable if and only if T^* is diagonalizable.

7A.26 Fix $u, x \in V$. Define $T \in \text{End}(V)$ by $Tv = \langle v, u\rangle x$ for every $v \in V$.

(a) Prove that if V is a real vector space, then T is self-adjoint if and only if the list u, x is linearly dependent.

(b) Prove that T is normal if and only if the list u, x is linearly dependent.

7A.27 Suppose $T \in \text{End}(V)$ is normal. Prove that

$$\ker T^k = \ker T \quad \text{and} \quad \text{im}\, T^k = \text{im}\, T$$

for every positive integer k.

7A.28 Suppose $T \in \text{End}(V)$ is normal. Prove that if $\lambda \in \mathbb{F}$, then the minimal polynomial of T is not a polynomial multiple of $(x - \lambda)^2$.

7A.29 Prove or give a counterexample: If $T \in \text{End}(V)$ and there is an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Te_k\| = \|T^*e_k\|$ for each $k = 1, \dots, n$, then T is normal.

7A.30 Suppose that $T \in \text{End}(\mathbb{F}^3)$ is normal and $T(1, 1, 1) = (2, 2, 2)$. Suppose $(z_1, z_2, z_3) \in \ker T$. Prove that $z_1 + z_2 + z_3 = 0$.

7A.31 Fix a positive integer n. In the inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product $\langle f, g\rangle = \int_{-\pi}^{\pi} fg$, let

$$V = \text{span}(1, \cos x, \cos 2x, \dots, \cos nx, \sin x, \sin 2x, \dots, \sin nx).$$

(a) Define $D \in \text{End}(V)$ by $Df = f'$. Show that $D^* = -D$. Conclude that D is normal but not self-adjoint.

(b) Define $T \in \text{End}(V)$ by $Tf = f''$. Show that T is self-adjoint.

7A.32 Suppose $T: V \to W$ is a linear map. Show that under the standard identification of V with V' (see 6.62) and the corresponding identification of W with W', the adjoint map $T^*: W \to V$ corresponds to the dual map $T': W' \to V'$. More precisely, show that

$$T'(\varphi_w) = \varphi_{T^*w}$$

for all $w \in W$, where φ_w and φ_{T^*w} are defined as in 6.62.

7B Spectral Theorem

Recall that a diagonal matrix is a square matrix that is 0 everywhere except possibly on the diagonal. Recall that an operator on V is called diagonalizable if the operator has a diagonal matrix with respect to some basis of V. Recall also that this happens if and only if there is a basis of V consisting of eigenvectors of the operator (see 5.55).

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 $T \in \text{End}(V)$ 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 self-adjoint operators when $\mathbb{F} = \mathbb{R}$ and as the normal operators when $\mathbb{F} = \mathbb{C}$.

The spectral theorem is probably the most useful tool in the study of operators on inner product spaces. Its extension to certain infinite-dimensional inner product spaces (see, for example, Section 10D of the author's book *Measure, Integration & Real Analysis*) plays a key role in functional analysis.

Because the conclusion of the spectral theorem depends on $\mathbb{F}$, we will break the spectral theorem into two pieces, called the real spectral theorem and the complex spectral theorem.

Real Spectral Theorem

To prove the real spectral theorem, we will need two preliminary results. These preliminary results hold on both real and complex inner product spaces, but they are not needed for the proof of the complex spectral theorem.

You could guess that the next result is true and even discover its proof by thinking about quadratic polynomials with real coefficients. Specifically, suppose $b, c \in \mathbb{R}$ and $b^2 < 4c$. Let x be a real number. Then

This completing-the-square technique can be used to derive the quadratic formula.

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0.$$

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) leads to the next result.

Invertible quadratic expressions **Result 7.26**

Suppose $T \in \text{End}(V)$ is self-adjoint and $b, c \in \mathbb{R}$ are such that $b^2 < 4c$. Then

$$T^2 + bT + cI$$

is an invertible operator.

Proof. Let v be a nonzero vector in V. Then

$$\begin{aligned}\big\langle (T^2 + bT + cI)v, v\big\rangle &= \langle T^2 v, v\rangle + b\langle Tv, v\rangle + c\langle v, v\rangle \\ &= \langle Tv, Tv\rangle + b\langle Tv, v\rangle + c\|v\|^2 \\ &\geq \|Tv\|^2 - |b|\,\|Tv\|\,\|v\| + c\|v\|^2 \\ &= \Big(\|Tv\| - \frac{|b|\,\|v\|}{2}\Big)^2 + \Big(c - \frac{b^2}{4}\Big)\|v\|^2 \\ &> 0,\end{aligned}$$

where the third line above holds by the Cauchy–Schwarz inequality (6.14). 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 3.65). □

The next result will be a key tool in our proof of the real spectral theorem.

Minimal polynomial of self-adjoint operator **Result 7.27**

Suppose $T \in \text{End}(V)$ is self-adjoint. Then the minimal polynomial of T equals $(z - \lambda_1)\cdots(z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$.

Proof. First suppose $\mathbb{F} = \mathbb{C}$. The zeros of the minimal polynomial of T are the eigenvalues of T [by 5.27(a)]. All eigenvalues of T are real (by 7.12). Thus the second version of the fundamental theorem of algebra (see 4.13) tells us that the minimal polynomial of T has the desired form.

Now suppose $\mathbb{F} = \mathbb{R}$. By the factorization of a polynomial over $\mathbb{R}$ (see 4.16) there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ and $b_1, \ldots, b_N, c_1, \ldots, c_N \in \mathbb{R}$ with $b_k^2 < 4c_k$ for each k such that the minimal polynomial of T equals

$$(z - \lambda_1)\cdots(z - \lambda_m)(z^2 + b_1 z + c_1)\cdots(z^2 + b_N z + c_N); \tag{7.28}$$

here either m or N might equal 0, meaning that there are no terms of the corresponding form. Now

$$(T - \lambda_1 I)\cdots(T - \lambda_m I)(T^2 + b_1 T + c_1 I)\cdots(T^2 + b_N T + c_N I) = 0.$$

If $N > 0$, then we could multiply both sides of the equation above on the right by the inverse of $T^2 + b_N T + c_N I$ (which is an invertible operator by 7.26) to obtain a polynomial expression of T that equals 0. The corresponding polynomial would have degree two less than the degree of 7.28, violating the minimality of the degree of the polynomial with this property. Thus we must have $N = 0$, which means that the minimal polynomial in 7.28 has the form $(z - \lambda_1)\cdots(z - \lambda_m)$, as desired. □

The result above along with 5.27(a) implies that every self-adjoint operator has an eigenvalue. In fact, as we will see in the next result, self-adjoint operators have enough eigenvectors to form a basis.

The next result, which gives a complete description of the self-adjoint operators on a real inner product space, is one of the major theorems in linear algebra.

Real spectral theorem **Result 7.29**

Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \text{End}(V)$. Then the following are equivalent.

(a) T is self-adjoint.

(b) T has a diagonal matrix with respect to some orthonormal basis of V.

(c) V has an orthonormal basis consisting of eigenvectors of T.

Proof. First suppose (a) holds, so T is self-adjoint. Our results on minimal polynomials, specifically 6.41 and 7.27, imply that T has an upper-triangular matrix with respect to some orthonormal basis of V. With respect to this orthonormal basis, the matrix of T^* is the transpose of the matrix of T. However, $T^* = T$. Thus the transpose of the matrix of T equals the matrix of T. Because the matrix of T is upper-triangular, this means that all entries of the matrix above and below the diagonal are 0. Hence the matrix of T is a diagonal matrix with respect to the orthonormal basis. Thus (a) implies (b).

Conversely, now suppose (b) holds, so T has a diagonal matrix with respect to some orthonormal basis of V. That diagonal matrix equals its transpose. Thus with respect to that basis, the matrix of T^* equals the matrix of T. Hence $T^* = T$, proving that (b) implies (a).

The equivalence of (b) and (c) follows from the definitions [or see the proof that (a) and (b) are equivalent in 5.55]. □

Example 7.30 **An orthonormal basis of eigenvectors for an operator**

Consider the operator T on $\mathbb{R}^3$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{pmatrix}.$$

This matrix with real entries equals its transpose; thus T is self-adjoint. As you can verify,

$$\frac{(1,-1,0)}{\sqrt{2}}, \quad \frac{(1,1,1)}{\sqrt{3}}, \quad \frac{(1,1,-2)}{\sqrt{6}}$$

is an orthonormal basis of $\mathbb{R}^3$ consisting of eigenvectors of T. With respect to this basis, the matrix of T is the diagonal matrix

$$\begin{pmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -15 \end{pmatrix}.$$

See Exercise 17 for a version of the real spectral theorem that applies simultaneously to more than one operator.

Complex Spectral Theorem

The next result gives a complete description of the normal operators on a complex inner product space.

Result 7.31

Complex spectral theorem
Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Then the following are equivalent.

(a) T is normal.

(b) T has a diagonal matrix with respect to some orthonormal basis of V.

(c) V has an orthonormal basis consisting of eigenvectors of T.

Proof. First suppose (a) holds, so T is normal. By Schur's theorem (6.42), there is an orthonormal basis $e_1, \dots, e_n$ of V with respect to which T has an upper-triangular matrix. Thus we can write

$$\text{Mat}\big(T, (e_1, \dots, e_n)\big) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}. \tag{7.32}$$

We will show that this matrix is actually a diagonal matrix.

We see from the matrix above that

$$\begin{aligned} \|Te_1\|^2 &= |a_{1,1}|^2, \\ \|T^*e_1\|^2 &= |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2. \end{aligned}$$

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 7.32, except possibly the first entry $a_{1,1}$, equal 0.

Now 7.32 implies

$$\|Te_2\|^2 = |a_{2,2}|^2$$

(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 7.32, except possibly the diagonal entry $a_{2,2}$, equal 0.

Continuing in this fashion, we see that all nondiagonal entries in the matrix 7.32 equal 0. Thus (b) holds, completing the proof that (a) implies (b).

Now suppose (b) 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, completing the proof that (b) implies (a).

The equivalence of (b) and (c) follows from the definitions (also see 5.55). □

See Exercises 13 and 20 for alternative proofs that (a) implies (b) in the previous result.

Exercises 14 and 15 interpret the real spectral theorem and the complex spectral theorem by expressing the domain space as an orthogonal direct sum of eigenspaces.

See Exercise 16 for a version of the complex spectral theorem that applies simultaneously to more than one operator.

The main conclusion of the complex spectral theorem is that every normal operator on a complex finite-dimensional inner product space is diagonalizable by an orthonormal basis, as illustrated by the next example.

Example 7.33 **An orthonormal basis of eigenvectors for an operator**

Consider the operator $T \in \text{End}(\mathbb{C}^2)$ defined by $T(w, z) = (2w - 3z, 3w + 2z)$. The matrix of T (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

As we saw in Example 7.19, T is a normal operator.

As you can verify,

$$(1/\sqrt{2})(i, 1),\ (1/\sqrt{2})(-i, 1)$$

is an orthonormal basis of $\mathbb{C}^2$ consisting of eigenvectors of T, and with respect to this basis the matrix of T is the diagonal matrix

$$\begin{pmatrix} 2 + 3i & 0 \\ 0 & 2 - 3i \end{pmatrix}.$$

Exercises

7B.1 Prove that a normal operator on a complex inner product space is self-adjoint if and only if all its eigenvalues are real.

This exercise strengthens the analogy (for normal operators) between self-adjoint operators and real numbers.

7B.2 Suppose $\mathbb{F} = \mathbb{C}$. Suppose $T \in \text{End}(V)$ is normal and has only one eigenvalue. Prove that T is a scalar multiple of the identity operator.

7B.3 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$ is normal. Prove that the set of eigenvalues of T is contained in $\{0, 1\}$ if and only if there is a subspace U of V such that $T = P_U$.

7B.4 Prove that a normal operator on a complex inner product space is skew (meaning it equals the negative of its adjoint) if and only if all its eigenvalues are purely imaginary (meaning that they have real part equal to 0).

7B.5 Prove or give a counterexample: If $T \in \text{End}(\mathbb{C}^3)$ is a diagonalizable operator, then T is normal (with respect to the usual inner product).

7B.6 Suppose V is a complex inner product space and $T \in \text{End}(V)$ is a normal operator such that $T^9 = T^8$. Prove that T is self-adjoint and $T^2 = T$.

7B.7 Give an example of an operator T on a complex vector space such that $T^9 = T^8$ but $T^2 \neq T$.

7B.8 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that T is normal if and only if every eigenvector of T is also an eigenvector of T^*.

7B.9 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that T is normal if and only if there exists a polynomial $p \in \mathbb{C}[X]$ such that $T^* = p(T)$.

7B.10 Suppose V is a complex inner product space. Prove that every normal operator on V has a square root.

An operator $S \in \text{End}(V)$ is called a ***square root*** *of $T \in \text{End}(V)$ if $S^2 = T$. We will discuss more about square roots of operators in Sections 7C and 8C.*

7B.11 Prove that every self-adjoint operator on V has a cube root.

An operator $S \in \text{End}(V)$ is called a ***cube root*** *of $T \in \text{End}(V)$ if $S^3 = T$.*

7B.12 Suppose V is a complex vector space and $T \in \text{End}(V)$ is normal. Prove that if S is an operator on V that commutes with T, then S commutes with T^*.

The result in this exercise is called Fuglede's theorem.

7B.13 Without using the complex spectral theorem, use the version of Schur's theorem that applies to two commuting operators (take $\mathscr{E} = \{T, T^*\}$ in Exercise 20 in Section 6B) to give a different proof that if $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$ is normal, then T has a diagonal matrix with respect to some orthonormal basis of V.

7B.14 Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \text{End}(V)$. 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.

7B.15 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. 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.

7B.16 Suppose $\mathbb{F} = \mathbb{C}$ and $\mathscr{E} \subseteq \text{End}(V)$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathscr{E}$ has a diagonal matrix if and only if S and T are commuting normal operators for all $S, T \in \mathscr{E}$.

This exercise extends the complex spectral theorem to the context of a collection of commuting normal operators.

7B.17 Suppose $\mathbb{F} = \mathbb{R}$ and $\mathscr{E} \subseteq \text{End}(V)$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathscr{E}$ has a diagonal matrix if and only if S and T are commuting self-adjoint operators for all $S, T \in \mathscr{E}$.

This exercise extends the real spectral theorem to the context of a collection of commuting self-adjoint operators.

7B.18 Give an example of a real inner product space V, an operator $T \in \text{End}(V)$, and real numbers b, c with $b^2 < 4c$ such that

$$T^2 + bT + cI$$

is not invertible.

This exercise shows that the hypothesis that T is self-adjoint cannot be deleted in 7.26, even for real vector spaces.

7B.19 Suppose $T \in \text{End}(V)$ is self-adjoint and U is a subspace of V that is invariant under T.

(a) Prove that $U^\perp$ is invariant under T.

(b) Prove that $T|_U \in \text{End}(U)$ is self-adjoint.

(c) Prove that $T|_{U^\perp} \in \text{End}(U^\perp)$ is self-adjoint.

7B.20 Suppose $T \in \text{End}(V)$ is normal and U is a subspace of V that is invariant under T.

(a) Prove that $U^\perp$ is invariant under T.

(b) Prove that U is invariant under T^*.

(c) Prove that $(T|_U)^* = (T^*)|_U$.

(d) Prove that $T|_U \in \text{End}(U)$ and $T|_{U^\perp} \in \text{End}(U^\perp)$ are normal operators.

This exercise can be used to give yet another proof of the complex spectral theorem (use induction on $\dim V$ *and the result that T has an eigenvector).*

7B.21 Suppose 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

$$T^2 - 5T + 6I = 0.$$

7B.22 Give an example of an operator $T \in \text{End}(\mathbb{C}^3)$ such that 2 and 3 are the only eigenvalues of T and $T^2 - 5T + 6I \neq 0$.

7B.23 Suppose $T \in \text{End}(V)$ is self-adjoint, $\lambda \in \mathbb{F}$, 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 $|\lambda - \lambda'| < \epsilon$.

This exercise shows that for a self-adjoint operator, a number that is close to satisfying an equation that would make it an eigenvalue is close to an eigenvalue.

7B.24 Suppose U is a finite-dimensional vector space and $T \in \text{End}(U)$.

(a) Suppose $\mathbb{F} = \mathbb{R}$. Prove that T is diagonalizable if and only if there is a basis of U such that the matrix of T with respect to this basis equals its transpose.

(b) Suppose $\mathbb{F} = \mathbb{C}$. Prove that T is diagonalizable if and only if there is a basis of U such that the matrix of T with respect to this basis commutes with its conjugate transpose.

This exercise adds another equivalence to the list of conditions equivalent to diagonalizability in 5.55.

7B.25 Suppose that $T \in \text{End}(V)$ and there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. Show that if $k \in \{1, \dots, n\}$, then the pseudoinverse $T^\dagger$ satisfies the equation

$$T^\dagger e_k = \begin{cases} \frac{1}{\lambda_k} e_k & \text{if } \lambda_k \neq 0, \\ 0 & \text{if } \lambda_k = 0. \end{cases}$$

7C Positive Operators

Positive operator — Definition 7.34

An operator $T \in \text{End}(V)$ 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 be self-adjoint can be dropped from the definition above (by 7.14).

Example 7.35 **Positive operators**

(a) Let $T \in \text{End}(\mathbb{F}^2)$ be the operator whose matrix (using the standard basis) is $\left(\begin{smallmatrix} 2 & -1 \\ -1 & 1 \end{smallmatrix}\right)$. Then T is self-adjoint and $\langle T(w,z),(w,z)\rangle = 2|w|^2 - 2\,\text{Re}(w\overline{z}) + |z|^2 = |w-z|^2 + |w|^2 \geq 0$ for all $(w,z) \in \mathbb{F}^2$. Thus T is a positive operator.

(b) If U is a subspace of V, then the orthogonal projection P_U is a positive operator, as you should verify.

(c) If $T \in \text{End}(V)$ is self-adjoint and $b, c \in \mathbb{R}$ are such that $b^2 < 4c$, then $T^2 + bT + cI$ is a positive operator, as shown by the proof of 7.26.

Square root **Definition 7.36**

An operator R is called a *square root* of an operator T if $R^2 = T$.

Example 7.37 **Square root of an operator**

If $T \in \text{End}(\mathbb{F}^3)$ is defined by $T(z_1,z_2,z_3) = (z_3,0,0)$, then the operator $R \in \text{End}(\mathbb{F}^3)$ defined by $R(z_1,z_2,z_3) = (z_2,z_3,0)$ is a square root of T because $R^2 = T$, as you can verify.

The characterizations of the positive operators in the next result correspond to characterizations of the nonnegative numbers among $\mathbb{C}$. Specifically, a number $z \in \mathbb{C}$ is nonnegative if and only if it has a nonnegative square root, corresponding to condition (d). Also, z is nonnegative if and only if it has a real square root, corresponding to condition (e). Finally, z is nonnegative if and only if there exists $w \in \mathbb{C}$ such that $z = \overline{w}w$, corresponding to condition (f). See Exercise 20 for another condition that is equivalent to being a positive operator.

Because positive operators correspond to nonnegative numbers, better terminology would use the term *nonnegative operators*. However, operator theorists consistently call these positive operators, so we follow that custom. Some mathematicians use the term **positive semidefinite operator**, which means the same as positive operator.

Characterizations of positive operators **Result 7.38**

Let $T \in \text{End}(V)$. Then the following are equivalent.

(a) T is a positive operator.

(b) T is self-adjoint and all eigenvalues of T are nonnegative.

(c) With respect to some orthonormal basis of V, the matrix of T is a diagonal matrix with only nonnegative numbers on the diagonal.

(d) T has a positive square root.

(e) T has a self-adjoint square root.

(f) $T = R^*R$ for some $R \in \text{End}(V)$.

Proof. We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (a).

First suppose (a) holds, so that T is positive, which implies that T is self-adjoint (by definition of 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

$$0 \le \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle.$$

Thus λ is a nonnegative number. Hence (b) holds, showing that (a) implies (b).

Now suppose (b) holds, so that T is self-adjoint and all eigenvalues of T are nonnegative. By the spectral theorem (7.29 and 7.31), 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_k is a nonnegative number. The matrix of T with respect to $e_1, \dots, e_n$ is the diagonal matrix with $\lambda_1, \dots, \lambda_n$ on the diagonal, which shows that (b) implies (c).

Now suppose (c) holds. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V such that the matrix of T with respect to this basis is a diagonal matrix with nonnegative numbers $\lambda_1, \dots, \lambda_n$ on the diagonal. The linear map lemma (3.4) implies that there exists $R \in \text{End}(V)$ such that

$$Re_k = \sqrt{\lambda_k}\, e_k$$

for each $k = 1, \dots, n$. As you should verify, R is a positive operator. Furthermore, $R^2 e_k = \lambda_k e_k = Te_k$ for each k, which implies that $R^2 = T$. Thus R is a positive square root of T. Hence (d) holds, which shows that (c) implies (d).

Every positive operator is self-adjoint (by definition of positive operator). Thus (d) implies (e).

Now suppose (e) 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) implies (f).

Finally, suppose (f) holds. Let $R \in \text{End}(V)$ 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

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle \ge 0$$

for every $v \in V$. Thus T is positive, showing that (f) implies (a). □

Every nonnegative number has a unique nonnegative square root. The next result shows that positive operators enjoy a similar property.

Each positive operator has only one positive square root **Result 7.39**

Every positive operator on V has a unique positive square root.

Proof. Suppose $T \in \text{End}(V)$ is positive. Suppose $v \in V$ is an eigenvector of T. Hence there exists a real number $\lambda \ge 0$ such that $Tv = \lambda v$.

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

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 $\dim V > 1$.

positive operator, all its eigenvalues are nonnegative. Thus there exist nonnegative numbers $\lambda_1, \dots, \lambda_n$ such that $Re_k = \sqrt{\lambda_k}\, e_k$ for each $k = 1, \dots, n$.

Because $e_1, \dots, e_n$ is a basis of V, we can write

$$v = a_1 e_1 + \cdots + a_n e_n$$

for some numbers $a_1, \dots, a_n \in \mathbb{F}$. Thus

$$Rv = a_1 \sqrt{\lambda_1}\, e_1 + \cdots + a_n \sqrt{\lambda_n}\, e_n.$$

Hence

$$\lambda v = Tv = R^2 v = a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n.$$

The equation above implies that

$$a_1 \lambda e_1 + \cdots + a_n \lambda e_n = a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n.$$

Thus $a_k(\lambda - \lambda_k) = 0$ for each $k = 1, \dots, n$. Hence

$$v = \sum_{\{k : \lambda_k = \lambda\}} a_k e_k.$$

Thus

$$Rv = \sum_{\{k : \lambda_k = \lambda\}} a_k \sqrt{\lambda}\, e_k = \sqrt{\lambda}\, v,$$

as desired. □

The notation defined below makes sense thanks to the result above.

Notation 7.40

$\sqrt{T}$

For T a positive operator, $\sqrt{T}$ denotes the unique positive square root of T.

Example 7.41 **Square root of positive operators**

Define operators S, T on $\mathbb{R}^2$ (with the usual Euclidean inner product) by

$$S(x, y) = (x, 2y) \quad \text{and} \quad T(x, y) = (x + y, x + y).$$

Then with respect to the standard basis of $\mathbb{R}^2$ we have

$$\mathrm{Mat}(S) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathrm{Mat}(T) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{7.42}$$

Each of these matrices equals its transpose; thus S and T are self-adjoint.

If $(x, y) \in \mathbb{R}^2$, then

$$\langle S(x, y), (x, y) \rangle = x^2 + 2y^2 \geq 0$$

and

$$\langle T(x, y), (x, y) \rangle = x^2 + 2xy + y^2 = (x + y)^2 \geq 0.$$

Example 7.41
continued

Thus S and T are positive operators.

The standard basis of $\mathbb{R}^2$ is an orthonormal basis consisting of eigenvectors of S. Note that

$$\big(1/\sqrt{2}, 1/\sqrt{2}\big), \big(1/\sqrt{2}, -1/\sqrt{2}\big)$$

is an orthonormal basis of eigenvectors of T, with eigenvalue 2 for the first eigenvector and eigenvalue 0 for the second eigenvector. Thus $\sqrt{T}$ has the same eigenvectors, with eigenvalues $\sqrt{2}$ and 0.

You can verify that

$$\operatorname{Mat}\big(\sqrt{S}\big) = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{pmatrix} \quad \text{and} \quad \operatorname{Mat}\big(\sqrt{T}\big) = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

with respect to the standard basis by showing that the squares of the matrices above are the matrices in 7.42 and that each matrix above is the matrix of a positive operator.

The statement of the next result does not involve a square root, but the clean proof makes nice use of the square root of a positive operator.

T positive and $\langle Tv, v\rangle = 0 \implies Tv = 0$ **Result 7.43**
Suppose T is a positive operator on V and $v \in V$ is such that $\langle Tv, v\rangle = 0$. Then $Tv = 0$.

Proof. We have

$$0 = \langle Tv, v\rangle = \big\langle \sqrt{T}\sqrt{T}v, v\big\rangle = \big\langle \sqrt{T}v, \sqrt{T}v\big\rangle = \big\|\sqrt{T}v\big\|^2.$$

Hence $\sqrt{T}v = 0$. Thus $Tv = \sqrt{T}\big(\sqrt{T}v\big) = 0$, as desired. □

Exercises

7C.1 Suppose $T \in \operatorname{End}(V)$. Prove that if both T and $-T$ are positive operators, then $T = 0$.

7C.2 Suppose $T \in \operatorname{End}(\mathbb{F}^4)$ is the operator whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}.$$

Show that T is an invertible positive operator.

7C.3 Suppose n is a positive integer and $T \in \text{End}(\mathbb{F}^n)$ is the operator whose matrix (with respect to the standard basis) consists of all 1's. Show that T is a positive operator.

7C.4 Suppose n is an integer with $n > 1$. Show that there exists an n-by-n matrix A such that all of the entries of A are positive numbers and $A = A^*$, but the operator on $\mathbb{F}^n$ whose matrix (with respect to the standard basis) equals A is not a positive operator.

7C.5 Suppose $T \in \text{End}(V)$ is self-adjoint. Prove that T is a positive operator if and only if for every orthonormal basis $e_1, \dots, e_n$ of V, all entries on the diagonal of $\text{Mat}\big(T, (e_1, \dots, e_n)\big)$ are nonnegative numbers.

7C.6 Prove that the sum of two positive operators on V is a positive operator.

7C.7 Suppose $S \in \text{End}(V)$ is an invertible positive operator and $T \in \text{End}(V)$ is a positive operator. Prove that $S + T$ is invertible.

7C.8 Suppose $T \in \text{End}(V)$. Prove that T is a positive operator if and only if the pseudoinverse $T^\dagger$ is a positive operator.

7C.9 Suppose $T \in \text{End}(V)$ is a positive operator and $S \in \text{Hom}(W, V)$. Prove that S^*TS is a positive operator on W.

7C.10 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 $v = w$.

7C.11 Suppose T is a positive operator on V and U is a subspace of V invariant under T. Prove that $T|_U \in \text{End}(U)$ is a positive operator on U.

7C.12 Suppose $T \in \text{End}(V)$ is a positive operator. Prove that T^k is a positive operator for every positive integer k.

7C.13 Suppose $T \in \text{End}(V)$ is self-adjoint and $\alpha \in \mathbb{R}$.

(a) Prove that $T - \alpha I$ is a positive operator if and only if α is less than or equal to every eigenvalue of T.

(b) Prove that $\alpha I - T$ is a positive operator if and only if α is greater than or equal to every eigenvalue of T.

7C.14 Suppose T is a positive operator on V and $v_1, \dots, v_m \in V$. Prove that

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \langle T v_k, v_j \rangle \geq 0.$$

7C.15 Suppose $T \in \text{End}(V)$ is self-adjoint. Prove that there exist positive operators $A, B \in \text{End}(V)$ such that

$$T = A - B \quad \text{and} \quad \sqrt{T^* T} = A + B \quad \text{and} \quad AB = BA = 0.$$

7C.16 Suppose T is a positive operator on V. Prove that

$$\ker \sqrt{T} = \ker T \quad \text{and} \quad \text{im} \sqrt{T} = \text{im}\, T.$$

7C.17 Suppose that $T \in \text{End}(V)$ is a positive operator. Prove that there exists a polynomial p with real coefficients such that $\sqrt{T} = p(T)$.

7C.18 Suppose S and T are positive operators on V. Prove that ST is a positive operator if and only if S and T commute.

7C.19 Show that the identity operator on $\mathbb{F}^2$ has infinitely many self-adjoint square roots.

7C.20 Suppose $T \in \text{End}(V)$ and $e_1, \dots, e_n$ is an orthonormal basis of V. Prove that T is a positive operator if and only if there exist $v_1, \dots, v_n \in V$ such that

$$\langle T e_k, e_j \rangle = \langle v_k, v_j \rangle$$

for all $j, k = 1, \dots, n$.

The numbers $\{\langle T e_k, e_j \rangle\}_{j,k=1,\dots,n}$ are the entries in the matrix of T with respect to the orthonormal basis $e_1, \dots, e_n$.

7C.21 Suppose n is a positive integer. The n-by-n *Hilbert matrix* is the n-by-n matrix whose entry in row j, column k is $1/(j + k - 1)$. Suppose $T \in \text{End}(V)$ is an operator whose matrix with respect to some orthonormal basis of V is the n-by-n Hilbert matrix. Prove that T is a positive invertible operator.

Example: The 4-by-4 Hilbert matrix is

$$\begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{pmatrix}.$$

7C.22 Suppose $T \in \text{End}(V)$ is a positive operator and $u \in V$ is such that $\|u\| = 1$ and $\|Tu\| \geq \|Tv\|$ for all $v \in V$ with $\|v\| = 1$. Show that u is an eigenvector of T corresponding to the largest eigenvalue of T.

7C.23 For $T \in \text{End}(V)$ and $u, v \in V$, define $\langle u, v\rangle_T$ by $\langle u, v\rangle_T = \langle Tu, v\rangle$.

(a) Suppose $T \in \text{End}(V)$. 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 $\langle \cdot, \cdot \rangle$).

(b) Prove that every inner product on V is of the form $\langle \cdot, \cdot \rangle_T$ for some positive invertible operator $T \in \text{End}(V)$.

7C.24 Suppose S and T are positive operators on V. Prove that

$$\ker(S + T) = \ker S \cap \ker T.$$

7C.25 Let T be the second derivative operator in Exercise 31(b) in Section 7A. Show that $-T$ is a positive operator.

7D Isometries, Unitary Operators, and Matrix Factorization

Isometries

Linear maps that preserve norms are sufficiently important to deserve a name.

Definition 7.44

Isometry

A linear map $S \in \text{Hom}(V, W)$ is called an *isometry* if

$$\|Sv\| = \|v\|$$

for every $v \in V$. In other words, a linear map is an isometry if it preserves norms.

If $S \in \text{Hom}(V, W)$ is an isometry and $v \in V$ is such that $Sv = 0$, then

$$\|v\| = \|Sv\| = \|0\| = 0,$$

which implies that $v = 0$. Thus every isometry is injective.

The Greek word **isos** means equal; the Greek word **metron** means measure. Thus **isometry** literally means equal measure.

Example 7.45 **Orthonormal basis maps to orthonormal list $\Longrightarrow$ isometry**

Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $g_1, \ldots, g_n$ is an orthonormal list in W. Let $S \in \text{Hom}(V, W)$ be the linear map such that $Se_k = g_k$ for each $k = 1, \ldots, n$. To show that S is an isometry, suppose $v \in V$. Then

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n \tag{7.46}$$

Example 7.45
continued

and

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2, \tag{7.47}$$

where we have used 6.34(b). Applying S to both sides of 7.46 gives

$$Sv = \langle v, e_1\rangle Se_1 + \cdots + \langle v, e_n\rangle Se_n = \langle v, e_1\rangle g_1 + \cdots + \langle v, e_n\rangle g_n.$$

Thus

$$\|Sv\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2. \tag{7.48}$$

Comparing 7.47 and 7.48 shows that $\|v\| = \|Sv\|$. Thus S is an isometry.

The next result gives conditions equivalent to being an isometry. The equivalence of (a) and (c) shows that a linear map is an isometry if and only if it preserves inner products. The equivalence of (a) and (d) shows that a linear map is an isometry if and only if it maps some orthonormal basis to an orthonormal list. Thus the isometries given by Example 7.45 include all isometries. Furthermore, a linear map is an isometry if and only if it maps every orthonormal basis to an orthonormal list [because whether or not (a) holds does not depend on the basis $e_1, \ldots, e_n$].

The equivalence of (a) and (e) in the next result shows that a linear map is an isometry if and only if the columns of its matrix (with respect to any orthonormal bases) form an orthonormal list. Here we are identifying the columns of an m-by-n matrix with elements of $\mathbb{F}^m$ and then using the Euclidean inner product on $\mathbb{F}^m$.

Result 7.49

Characterizations of isometries

Suppose $S \in \text{Hom}(V, W)$. 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 the following are equivalent.

(a) S is an isometry.

(b) $S^*S = I$.

(c) $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$.

(d) $Se_1, \ldots, Se_n$ is an orthonormal list in W.

(e) The columns of $\text{Mat}\big(S, (e_1, \ldots, e_n), (f_1, \ldots, f_m)\big)$ form an orthonormal list in $\mathbb{F}^m$ with respect to the Euclidean inner product.

Proof. First suppose (a) holds, so S is an isometry. If $v \in V$ then

$$\big\langle (I - S^*S)v, v\big\rangle = \langle v, v\rangle - \langle S^*Sv, v\rangle = \|v\|^2 - \langle Sv, Sv\rangle = \|v\|^2 - \|Sv\|^2 = 0.$$

Hence the self-adjoint operator $I - S^*S$ equals 0 (by 7.16). Thus $S^*S = I$, proving that (a) implies (b).

Now suppose (b) holds, so $S^*S = I$. If $u, v \in V$ then

$$\langle Su, Sv\rangle = \langle S^*Su, v\rangle = \langle Iu, v\rangle = \langle u, v\rangle,$$

proving that (b) implies (c).

Now suppose that (c) holds, so $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$. Thus if $j, k \in \{1, \dots, n\}$, then

$$\langle Se_j, Se_k\rangle = \langle e_j, e_k\rangle.$$

Hence $Se_1, \dots, Se_n$ is an orthonormal list in W, proving that (c) implies (d).

Now suppose that (d) holds, so $Se_1, \dots, Se_n$ is an orthonormal list in W. Let $A = \operatorname{Mat}\big(S, (e_1, \dots, e_n), (f_1, \dots, f_m)\big)$. If $k, r \in \{1, \dots, n\}$, then

$$\sum_{j=1}^{m} A_{j,k}\overline{A_{j,r}} = \Big\langle \sum_{j=1}^{m} A_{j,k} f_j, \sum_{j=1}^{m} A_{j,r} f_j \Big\rangle = \langle Se_k, Se_r\rangle = \begin{cases} 1 & \text{if } k = r, \\ 0 & \text{if } k \neq r. \end{cases} \tag{7.50}$$

The left side of 7.50 is the inner product in $\mathbb{F}^m$ of columns k and r of A. Thus the columns of A form an orthonormal list in $\mathbb{F}^m$, proving that (d) implies (e).

Now suppose (e) holds, so the columns of the matrix A defined in the paragraph above form an orthonormal list in $\mathbb{F}^m$. Then 7.50 shows that $Se_1, \dots, Se_n$ is an orthonormal list in W. Thus Example 7.45, with $Se_1, \dots, Se_n$ playing the role of $g_1, \dots, g_n$, shows that S is an isometry, proving that (e) implies (a). □

As previously noted, every isometry is injective. Every injective operator on a finite-dimensional vector space is invertible (see 3.65). A standing assumption for this chapter is that V is a finite-dimensional inner product space. Thus we could delete the word "invertible" from the definition above without changing the meaning. The unnecessary word "invertible" has been retained in the definition above for consistency with the definition readers may encounter when learning about inner product spaces that are not necessarily finite-dimensional.

See Exercises 1 and 11 for additional conditions that are equivalent to being an isometry.

Although the words "unitary" and "isometry" mean the same thing for operators on finite-dimensional inner product spaces, remember that a unitary operator maps a vector space to itself, while an isometry maps a vector space to another (possibly different) vector space.

Unitary Operators

In this subsection, we confine our attention to linear maps from a vector space to itself. In other words, we will be working with operators.

Unitary operator — **Definition 7.51**

An operator $S \in \operatorname{End}(V)$ is called *unitary* if S is an invertible isometry.

Example 7.52 **Rotation of $\mathbb{R}^2$**

Suppose $\theta \in \mathbb{R}$ and S is the operator on $\mathbb{F}^2$ whose matrix with respect to the standard basis of $\mathbb{F}^2$ is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The two columns of this matrix form an orthonormal list in $\mathbb{F}^2$; hence S is an isometry [by the equivalence of (a) and (e) in 7.49]. Thus S is a unitary operator.

Example 7.52 *continued*

If $\mathbb{F} = \mathbb{R}$, then S is the operator of counterclockwise rotation by θ radians around the origin of $\mathbb{R}^2$. This observation gives us another way to think about why S is an isometry, because each rotation around the origin of $\mathbb{R}^2$ preserves norms.

The next result (7.53) lists several conditions that are equivalent to being a unitary operator. All the conditions equivalent to being an isometry in 7.49 should be added to this list. The extra conditions in 7.53 arise because of limiting the context to linear maps from a vector space to itself. For example, 7.49 shows that a linear map $S \in \text{Hom}(V, W)$ is an isometry if and only if $S^*S = I$, while 7.53 shows that an operator $S \in \text{End}(V)$ is a unitary operator if and only if $S^*S = SS^* = I$.

Another difference is that 7.49(d) mentions an orthonormal list, while 7.53(d) mentions an orthonormal basis. Also, 7.49(e) mentions the columns of $\text{Mat}(T)$, while 7.53(e) mentions the rows of $\text{Mat}(T)$. Furthermore, $\text{Mat}(T)$ in 7.49(e) is with respect to an orthonormal basis of V and an orthonormal basis of W, while $\text{Mat}(T)$ in 7.53(e) is with respect to a single basis of V doing double duty.

Result 7.53

Characterizations of unitary operators

Suppose $S \in \text{End}(V)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V. Then the following are equivalent.

(a) S is a unitary operator.

(b) $S^*S = SS^* = I$.

(c) S is invertible and $S^{-1} = S^*$.

(d) $Se_1, \dots, Se_n$ is an orthonormal basis of V.

(e) The rows of $\text{Mat}\big(S, (e_1, \dots, e_n)\big)$ form an orthonormal basis of $\mathbb{F}^n$ with respect to the Euclidean inner product.

(f) S^* is a unitary operator.

Proof. First suppose (a) holds, so S is a unitary operator. Hence

$$S^*S = I$$

by the equivalence of (a) and (b) in 7.49. Multiply both sides of this equation by S^{-1} on the right, getting $S^* = S^{-1}$. Thus $SS^* = SS^{-1} = I$, as desired, proving that (a) implies (b).

The definitions of invertible and inverse show that (b) implies (c).

Now suppose (c) holds, so S is invertible and $S^{-1} = S^*$. Thus $S^*S = I$. Hence $Se_1, \dots, Se_n$ is an orthonormal list in V, by the equivalence of (b) and (d) in 7.49. The length of this list equals $\dim V$. Thus $Se_1, \dots, Se_n$ is an orthonormal basis of V, proving that (c) implies (d).

Now suppose (d) holds, so $Se_1, \dots, Se_n$ is an orthonormal basis of V. The equivalence of (a) and (d) in 7.49 shows that S is a unitary operator. Thus

$$(S^*)^*S^* = SS^* = I,$$

where the last equation holds because we already showed that (a) implies (b) in this result. The equation above and the equivalence of (a) and (b) in 7.49 show that S^* is an isometry. Thus the columns of $\text{Mat}\big(S^*, (e_1, \dots, e_n)\big)$ form an orthonormal basis of $\mathbb{F}^n$ [by the equivalence of (a) and (e) of 7.49]. The rows of $\text{Mat}\big(S, (e_1, \dots, e_n)\big)$ are the complex conjugates of the columns of $\text{Mat}\big(S^*, (e_1, \dots, e_n)\big)$. Thus the rows of $\text{Mat}\big(S, (e_1, \dots, e_n)\big)$ form an orthonormal basis of $\mathbb{F}^n$, proving that (d) implies (e).

Now suppose (e) holds. Thus the columns of $\text{Mat}\big(S^*, (e_1, \dots, e_n)\big)$ form an orthonormal basis of $\mathbb{F}^n$. The equivalence of (a) and (e) in 7.49 shows that S^* is an isometry, proving that (e) implies (f).

Now suppose (f) holds, so S^* is a unitary operator. The chain of implications we have already proved in this result shows that (a) implies (f). Applying this result to S^* shows that $(S^*)^*$ is a unitary operator, proving that (f) implies (a).

We have shown that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (a), completing the proof. □

Recall our analogy between $\mathbb{C}$ and $\text{End}(V)$. Under this analogy, a complex number z corresponds to an operator $S \in \text{End}(V)$, and $\overline{z}$ corresponds to S^*. The real numbers ($z = \overline{z}$) correspond to the self-adjoint operators ($S = S^*$), and the nonnegative numbers correspond to the (badly named) positive operators.

Another distinguished subset of $\mathbb{C}$ is the unit circle, which consists of the complex numbers z such that $|z| = 1$. The condition $|z| = 1$ is equivalent to the condition $\overline{z}z = 1$. Under our analogy, this corresponds to the condition $S^*S = I$, which is equivalent to S being a unitary operator. Hence the analogy shows that the unit circle in $\mathbb{C}$ corresponds to the set of unitary operators. In the next two results, this analogy appears in the eigenvalues of unitary operators. Also see Exercise 15 for another example of this analogy.

Eigenvalues of unitary operators have absolute value 1 **Result 7.54**

Suppose λ is an eigenvalue of a unitary operator. Then $|\lambda| = 1$.

Proof. Suppose $S \in \text{End}(V)$ is a unitary operator and λ is an eigenvalue of S. Let $v \in V$ be such that $v \neq 0$ and $Sv = \lambda v$. Then

$$|\lambda| \, \|v\| = \|\lambda v\| = \|Sv\| = \|v\|.$$

Thus $|\lambda| = 1$, as desired. □

The next result characterizes unitary operators on finite-dimensional complex inner product spaces, using the complex spectral theorem as the main tool.

Description of unitary operators on complex inner product spaces **Result 7.55**

Suppose $\mathbb{F} = \mathbb{C}$ and $S \in \text{End}(V)$. Then the following are equivalent.

(a) S is a unitary operator.

(b) There is an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value 1.

Proof. Suppose (a) holds, so S is a unitary operator. The equivalence of (a) and (b) in 7.53 shows that S is normal. Thus the complex spectral theorem (7.31) shows that there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of S. Every eigenvalue of S has absolute value 1 (by 7.54), completing the proof that (a) implies (b).

Now suppose (b) holds. Let $e_1, \dots, e_n$ be an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ all have absolute value 1. Then $Se_1, \dots, Se_n$ is also an orthonormal basis of V because

$$\langle Se_j, Se_k \rangle = \langle \lambda_j e_j, \lambda_k e_k \rangle = \lambda_j \overline{\lambda_k} \langle e_j, e_k \rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k \end{cases}$$

for all $j, k = 1, \dots, n$. Thus the equivalence of (a) and (d) in 7.53 shows that S is unitary, proving that (b) implies (a). □

QR Factorization

In this subsection, we shift our attention from operators to matrices. This switch should give you good practice in identifying an operator with a square matrix (after picking a basis of the vector space on which the operator is defined). You should also become more comfortable with translating concepts and results back and forth between the context of operators and the context of square matrices.

When starting with n-by-n matrices instead of operators, unless otherwise specified assume that the associated operators live on $\mathbb{F}^n$ (with the Euclidean inner product) and that their matrices are computed with respect to the standard basis of $\mathbb{F}^n$.

We begin by making the following definition, transferring the notion of a unitary operator to a unitary matrix.

Unitary matrix — Definition 7.56

An n-by-n matrix is called *unitary* if its columns form an orthonormal list in $\mathbb{F}^n$.

In the definition above, we could have replaced "orthonormal list in $\mathbb{F}^n$" with "orthonormal basis of $\mathbb{F}^n$" because every orthonormal list of length n in an n-dimensional inner product space is an orthonormal basis. If $S \in \text{End}(V)$ and $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are orthonormal bases of V, then S is a unitary operator if and only if $\text{Mat}\big(S, (e_1, \dots, e_n), (f_1, \dots, f_n)\big)$ is a unitary matrix, as shown by the equivalence of (a) and (e) in 7.49. Also note that we could also have replaced "columns" in the definition above with "rows" by using the equivalence between conditions (a) and (e) in 7.53.

The next result, whose proof will be left as an exercise for the reader, gives some equivalent conditions for a square matrix to be unitary. In (c), Qv denotes the matrix product of Q and v, identifying elements of $\mathbb{F}^n$ with n-by-1 matrices (sometimes called column vectors). The norm in (c) below is the usual Euclidean norm on $\mathbb{F}^n$ that comes from the Euclidean inner product. In (d), Q^* denotes the conjugate transpose of the matrix Q, which corresponds to the adjoint of the associated operator.

Characterizations of unitary matrices **Result 7.57**

Suppose Q is an n-by-n matrix. Then the following are equivalent.

(a) Q is a unitary matrix.

(b) The rows of Q form an orthonormal list in $\mathbb{F}^n$.

(c) $\|Qv\| = \|v\|$ for every $v \in \mathbb{F}^n$.

(d) $Q^*Q = QQ^* = I$, the n-by-n matrix with 1's on the diagonal and 0's elsewhere.

The QR factorization stated and proved below is the main tool in the widely used QR algorithm (not discussed here) for finding good approximations to eigenvalues and eigenvectors of square matrices. In the result below, if the matrix A is in $\mathbb{F}^{n,n}$, then the matrices Q and R are also in $\mathbb{F}^{n,n}$.

QR factorization **Result 7.58**

Suppose A is a square matrix with linearly independent columns. Then there exist unique matrices Q and R such that Q is unitary, R is upper triangular with only positive numbers on its diagonal, and

$$A = QR.$$

Proof. Let $v_1, \dots, v_n$ denote the columns of A, thought of as elements of $\mathbb{F}^n$. Apply the Gram–Schmidt procedure (6.36) to the list $v_1, \dots, v_n$, getting an orthonormal basis $e_1, \dots, e_n$ of $\mathbb{F}^n$ such that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k) \tag{7.59}$$

for each $k = 1, \dots, n$. Let R be the n-by-n matrix defined by

$$R_{j,k} = \langle v_k, e_j \rangle,$$

where $R_{j,k}$ denotes the entry in row j, column k of R. If $j > k$, then e_j is orthogonal to $\operatorname{span}(e_1, \dots, e_k)$ and hence e_j is orthogonal to v_k (by 7.59). In other words, if $j > k$ then $\langle v_k, e_j \rangle = 0$. Thus R is an upper-triangular matrix.

Let Q be the unitary matrix whose columns are $e_1, \dots, e_n$. If $k \in \{1, \dots, n\}$, then the kth column of QR equals a linear combination of the columns of Q, with the coefficients for the linear combination coming from the kth column of R—see 3.51(a). Hence the kth column of QR equals

$$\langle v_k, e_1 \rangle e_1 + \cdots + \langle v_k, e_k \rangle e_k,$$

which equals v_k [by 6.34(a)], the kth column of A. Thus $A = QR$, as desired.

The equations defining the Gram–Schmidt procedure (see 6.36) show that each v_k equals a positive multiple of e_k plus a linear combination of $e_1, \dots, e_{k-1}$. Thus each $\langle v_k, e_k \rangle$ is a positive number. Hence all entries on the diagonal of R are positive numbers, as desired.

Finally, to show that Q and R are unique, suppose we also have $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ is unitary and $\widehat{R}$ is upper triangular with only positive numbers on its diagonal. Let $q_1, \dots, q_n$ denote the columns of $\widehat{Q}$. Thinking of matrix multiplication as above, we see that each v_k is a linear combination of $q_1, \dots, q_k$, with the coefficients coming from the kth column of $\widehat{R}$. This implies that $\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(q_1, \dots, q_k)$ and $\langle v_k, q_k \rangle > 0$. The uniqueness of the orthonormal lists satisfying these conditions (see Exercise 10 in Section 6B) now shows that $q_k = e_k$ for each $k = 1, \dots, n$. Hence $\widehat{Q} = Q$, which then implies that $\widehat{R} = R$, completing the proof of uniqueness. □

The proof of the QR factorization shows that the columns of the unitary matrix can be computed by applying the Gram–Schmidt procedure to the columns of the matrix to be factored. The next example illustrates the computation of the QR factorization based on the proof that we just completed.

Example 7.60 **QR factorization of a 3-by-3 matrix**

To find the QR factorization of the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -4 \\ 0 & 3 & 2 \end{pmatrix},$$

follow the proof of 7.58. Thus set v_1, v_2, v_3 equal to the columns of A:

$$v_1 = (1, 0, 0), \quad v_2 = (2, 1, 3), \quad v_3 = (1, -4, 2).$$

Apply the Gram–Schmidt procedure to v_1, v_2, v_3, producing the orthonormal list

$$e_1 = (1, 0, 0), \quad e_2 = \bigl(0, 1/\sqrt{10}, 3/\sqrt{10}\bigr), \quad e_3 = \bigl(0, -3/\sqrt{10}, 1/\sqrt{10}\bigr).$$

Still following the proof of 7.58, let Q be the unitary matrix whose columns are e_1, e_2, e_3:

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{10} & -3/\sqrt{10} \\ 0 & 3/\sqrt{10} & 1/\sqrt{10} \end{pmatrix}.$$

As in the proof of 7.58, let R be the 3-by-3 matrix whose entry in row j, column k is $\langle v_k, e_j \rangle$, which gives

$$R = \begin{pmatrix} 1 & 2 & 1 \\ 0 & \sqrt{10} & \sqrt{10}/5 \\ 0 & 0 & 7\sqrt{10}/5 \end{pmatrix}.$$

Note that R is indeed an upper-triangular matrix with only positive numbers on the diagonal, as required by the QR factorization.

Example 7.60 *continued*

Now matrix multiplication can verify that $A = QR$ is the desired factorization of A:

$$QR = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{10} & -3/\sqrt{10} \\ 0 & 3/\sqrt{10} & 1/\sqrt{10} \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & \sqrt{10} & \sqrt{10}/5 \\ 0 & 0 & 7\sqrt{10}/5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -4 \\ 0 & 3 & 2 \end{pmatrix} = A.$$

Thus $A = QR$, as expected.

The QR factorization will be the major tool used in the proof of the Cholesky factorization (7.63) in the next subsection. For another nice application of the QR factorization, see the proof of Hadamard's inequality (9.66).

If a QR factorization is available, then it can be used to solve a corresponding system of linear equations without using Gaussian elimination. Specifically, suppose A is an n-by-n square matrix with linearly independent columns. Suppose that $b \in \mathbb{F}^n$ and we want to solve the equation $Ax = b$ for $x = (x_1, \dots, x_n) \in \mathbb{F}^n$ (as usual, we are identifying elements of $\mathbb{F}^n$ with n-by-1 column vectors).

Suppose $A = QR$, where Q is unitary and R is upper triangular with only positive numbers on its diagonal (Q and R are computable from A using just the Gram–Schmidt procedure, as shown in the proof of 7.58). The equation $Ax = b$ is equivalent to the equation $QRx = b$. Multiplying both sides of this last equation by Q^* on the left and using 7.57(d) gives the equation

$$Rx = Q^*b.$$

The matrix Q^* is the conjugate transpose of the matrix Q. Thus computing Q^*b is straightforward. Because R is an upper-triangular matrix with positive numbers on its diagonal, the system of linear equations represented by the equation above can quickly be solved by first solving for x_n, then for x_{n-1}, and so on.

Cholesky Factorization

We begin this subsection with a characterization of positive invertible operators in terms of inner products.

Positive invertible operator Result 7.61

A self-adjoint operator $T \in \text{End}(V)$ is a positive invertible operator if and only if $\langle Tv, v\rangle > 0$ for every nonzero $v \in V$.

Proof. First suppose T is a positive invertible operator. If $v \in V$ and $v \neq 0$, then because T is invertible we have $Tv \neq 0$. This implies that $\langle Tv, v\rangle \neq 0$ (by 7.43). Hence $\langle Tv, v\rangle > 0$.

To prove the implication in the other direction, suppose now that $\langle Tv, v\rangle > 0$ for every nonzero $v \in V$. Thus $Tv \neq 0$ for every nonzero $v \in V$. Hence T is injective. Thus T is invertible, as desired. □

The next definition transfers the result above to the language of matrices. Here we are using the usual Euclidean inner product on $\mathbb{F}^n$ and identifying elements of $\mathbb{F}^n$ with n-by-1 column vectors.

Positive definite — Definition 7.62

A matrix $B \in \mathbb{F}^{n,n}$ is called *positive definite* if $B^* = B$ and

$$\langle Bx, x\rangle > 0$$

for every nonzero $x \in \mathbb{F}^n$.

A matrix is upper triangular if and only if its conjugate transpose is lower triangular (meaning that all entries above the diagonal are 0). The factorization below, which has important consequences in computational linear algebra, writes a positive definite matrix as the product of a lower triangular matrix and its conjugate transpose.

Our next result is solely about matrices, although the proof makes use of the identification of results about operators with results about square matrices. In the result below, if the matrix B is in $\mathbb{F}^{n,n}$, then the matrix R is also in $\mathbb{F}^{n,n}$.

Cholesky factorization — Result 7.63

Suppose B is a positive definite matrix. Then there exists a unique upper-triangular matrix R with only positive numbers on its diagonal such that

$$B = R^*R.$$

Proof. Because B is positive definite, there exists an invertible square matrix A of the same size as B such that $B = A^*A$ [by the equivalence of (a) and (f) in 7.38].

Let $A = QR$ be the QR factorization of A (see 7.58), where Q is unitary and R is upper triangular with only positive numbers on its diagonal. Then $A^* = R^*Q^*$.

Thus

$$B = A^*A = R^*Q^*QR = R^*R,$$

as desired.

André-Louis Cholesky (1875–1918) discovered this factorization, which was published posthumously in 1924.

To prove the uniqueness part of this result, suppose S is an upper-triangular matrix with only positive numbers on its diagonal and $B = S^*S$. The matrix S is invertible because B is invertible (see Exercise 11 in Section 3D). Multiplying both sides of the equation $B = S^*S$ by S^{-1} on the right gives the equation $BS^{-1} = S^*$.

Let A be the matrix from the first paragraph of this proof. Then

$$\begin{aligned}(AS^{-1})^*(AS^{-1}) &= (S^*)^{-1}A^*AS^{-1}\\ &= (S^*)^{-1}BS^{-1}\\ &= (S^*)^{-1}S^*\\ &= I.\end{aligned}$$

Thus AS^{-1} is unitary.

Hence $A = (AS^{-1})S$ is a factorization of A as the product of a unitary matrix and an upper-triangular matrix with only positive numbers on its diagonal. The uniqueness of the QR factorization, as stated in 7.58, now implies that $S = R$. □

In the first paragraph of the proof above, we could have chosen A to be the unique positive definite matrix that is a square root of B (see 7.39). However, the proof was presented with the more general choice of A because for specific positive definite matrices B, it may be easier to find a different choice of A.

Exercises

7D.1 Suppose $\dim V \geq 2$ and $S \in \text{Hom}(V, W)$. Prove that S is an isometry if and only if Se_1, Se_2 is an orthonormal list in W for every orthonormal list e_1, e_2 of length two in V.

7D.2 Suppose $T \in \text{Hom}(V, W)$ and $T \neq 0$. Prove that T is a scalar multiple of an isometry if and only if T preserves orthogonality.

The phrase "T preserves orthogonality" means that $\langle Tu, Tv\rangle = 0$ for all $u, v \in V$ such that $\langle u, v\rangle = 0$.

7D.3

(a) Show that the product of two unitary operators on V is a unitary operator.

(b) Show that the inverse of a unitary operator on V is a unitary operator.

This exercise shows that the set of unitary operators on V is a group, where the group operation is the usual product of two operators.

7D.4 Suppose $\mathbb{F} = \mathbb{C}$ and $A, B \in \text{End}(V)$ are self-adjoint. Show that $A + iB$ is unitary if and only if $AB = BA$ and $A^2 + B^2 = I$.

7D.5 Suppose $S \in \text{End}(V)$. Prove that the following are equivalent.

(a) S is a self-adjoint unitary operator.

(b) $S = 2P - I$ for some orthogonal projection P on V.

(c) There exists a subspace U of V such that $Su = u$ for every $u \in U$ and $Sw = -w$ for every $w \in U^{\perp}$.

7D.6 Suppose T_1, T_2 are both normal operators on $\mathbb{F}^3$ with $2, 5, 7$ as eigenvalues. Prove that there exists a unitary operator $S \in \text{End}(\mathbb{F}^3)$ such that $T_1 = S^* T_2 S$.

7D.7 Give an example of two self-adjoint operators $T_1, T_2 \in \text{End}(\mathbb{F}^4)$ such that the eigenvalues of both operators are $2, 5, 7$ but there does not exist a unitary operator $S \in \text{End}(\mathbb{F}^4)$ such that $T_1 = S^* T_2 S$. Be sure to explain why there is no unitary operator with the required property.

7D.8 Prove or give a counterexample: If $S \in \text{End}(V)$ and there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $\|Se_k\| = 1$ for each e_k, then S is a unitary operator.

7D.9 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Suppose every eigenvalue of T has absolute value 1 and $\|Tv\| \le \|v\|$ for every $v \in V$. Prove that T is a unitary operator.

7D.10 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$ is a self-adjoint operator such that $\|Tv\| \le \|v\|$ for all $v \in V$.

(a) Show that $I - T^2$ is a positive operator.

(b) Show that $T + i\sqrt{I - T^2}$ is a unitary operator.

7D.11 Suppose $S \in \text{End}(V)$. Prove that S is a unitary operator if and only if

$$\{Sv : v \in V \text{ and } \|v\| \le 1\} = \{v \in V : \|v\| \le 1\}.$$

7D.12 Prove or give a counterexample: If $S \in \text{End}(V)$ is invertible and $\|S^{-1}v\| = \|Sv\|$ for every $v \in V$, then S is unitary.

7D.13 Explain why the columns of a square matrix of complex numbers form an orthonormal list in $\mathbb{C}^n$ if and only if the rows of the matrix form an orthonormal list in $\mathbb{C}^n$.

7D.14 Suppose $v \in V$ with $\|v\| = 1$ and $b \in \mathbb{F}$. Also suppose $\dim V \ge 2$. Prove that there exists a unitary operator $S \in \text{End}(V)$ such that $\langle Sv, v\rangle = b$ if and only if $|b| \le 1$.

7D.15 Suppose T is a unitary operator on V such that $T - I$ is invertible.

(a) Prove that $(T+I)(T-I)^{-1}$ is a skew operator (meaning that it equals the negative of its adjoint).

(b) Prove that if $\mathbb{F} = \mathbb{C}$, then $i(T + I)(T - I)^{-1}$ is a self-adjoint operator.

The function $z \mapsto i(z+1)(z-1)^{-1}$ maps the unit circle in $\mathbb{C}$ (except for the point 1) to $\mathbb{R}$. Thus (b) illustrates the analogy between the unitary operators and the unit circle in $\mathbb{C}$, along with the analogy between the self-adjoint operators and $\mathbb{R}$.

7D.16 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$ is self-adjoint. Prove that $(T + iI)(T - iI)^{-1}$ is a unitary operator and 1 is not an eigenvalue of this operator.

7D.17 Explain why the characterizations of unitary matrices given by 7.57 hold.

7D.18 A square matrix A is called *symmetric* if it equals its transpose. Prove that if A is a symmetric matrix with real entries, then there exists a unitary matrix Q with real entries such that Q^*AQ is a diagonal matrix.

7D.19 Suppose n is a positive integer. For this exercise, we adopt the notation that a typical element z of $\mathbb{C}^n$ is denoted by $z = (z_0, z_1, \ldots, z_{n-1})$. Define linear functionals $\omega_0, \omega_1, \ldots, \omega_{n-1}$ on $\mathbb{C}^n$ by

$$\omega_j(z_0, z_1, \ldots, z_{n-1}) = \frac{1}{\sqrt{n}} \sum_{m=0}^{n-1} z_m e^{-2\pi i j m/n}.$$

The *discrete Fourier transform* is the operator $\mathcal{F}: \mathbb{C}^n \to \mathbb{C}^n$ defined by

$$\mathcal{F}z = \big(\omega_0(z), \omega_1(z), \ldots, \omega_{n-1}(z)\big).$$

(a) Show that $\mathcal{F}$ is a unitary operator on $\mathbb{C}^n$.

(b) Show that if $(z_0, \ldots, z_{n-1}) \in \mathbb{C}^n$ and z_n is defined to equal z_0, then

$$\mathcal{F}^{-1}(z_0, z_1, \ldots, z_{n-1}) = \mathcal{F}(z_n, z_{n-1}, \ldots, z_1).$$

(c) Show that $\mathcal{F}^4 = I$.

The discrete Fourier transform has many important applications in data analysis. The usual Fourier transform involves expressions of the form $\int_{-\infty}^{\infty} f(x)e^{-2\pi i t x}\, dx$ for complex-valued integrable functions f defined on $\mathbb{R}$.

7D.20 Suppose A is a square matrix with linearly independent columns. Prove that there exist unique matrices R and Q such that R is lower triangular with only positive numbers on its diagonal, Q is unitary, and $A = RQ$.

7E Singular Value Decomposition

Singular Values

We will need the following result in this section.

Properties of T^*T **Result 7.64**

Suppose $T \in \text{Hom}(V, W)$. Then

(a) T^*T is a positive operator on V;

(b) $\ker T^*T = \ker T$;

(c) $\text{im}\, T^*T = \text{im}\, T^*$;

(d) $\dim \text{im}\, T = \dim \text{im}\, T^* = \dim \text{im}\, T^*T$.

Proof.

(a) We have

$$(T^*T)^* = T^*(T^*)^* = T^*T.$$

Thus T^*T is self-adjoint.

If $v \in V$, then

$$\big\langle (T^*T)v, v \big\rangle = \langle T^*(Tv), v \rangle = \langle Tv, Tv \rangle = \|Tv\|^2 \geq 0.$$

Thus T^*T is a positive operator.

(b) First suppose $v \in \ker T^*T$. Then

$$\|Tv\|^2 = \langle Tv, Tv \rangle = \langle T^*Tv, v \rangle = \langle 0, v \rangle = 0.$$

Thus $Tv = 0$, proving that $\ker T^*T \subseteq \ker T$.

The inclusion in the other direction is clear, because if $v \in V$ and $Tv = 0$, then $T^*Tv = 0$.

Thus $\ker T^*T = \ker T$, completing the proof of (b).

(c) We already know from (a) that T^*T is self-adjoint. Thus

$$\operatorname{im} T^*T = (\ker T^*T)^\perp = (\ker T)^\perp = \operatorname{im} T^*,$$

where the first and last equalities come from 7.6 and the second equality comes from (b).

(d) To verify the first equation in (d), note that

$$\dim \operatorname{im} T = \dim(\ker T^*)^\perp = \dim W - \dim \ker T^* = \dim \operatorname{im} T^*,$$

where the first equality comes from 7.6(d), the second equality comes from 6.55, and the last equality comes from the fundamental theorem of linear maps (3.21).

The equality $\dim \operatorname{im} T^* = \dim \operatorname{im} T^*T$ follows from (c). □

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. Eigenspaces and the notation E (used in the examples) were defined in 5.52.

Singular values Definition 7.65

Suppose $T \in \operatorname{Hom}(V, W)$. The *singular values* of T are the nonnegative square roots of the eigenvalues of T^*T, listed in decreasing order, each included as many times as the dimension of the corresponding eigenspace of T^*T.

Example 7.66 Singular values of an operator on $\mathbb{F}^4$

Define $T \in \operatorname{End}(\mathbb{F}^4)$ by $T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4)$. A calculation shows that

$$T^*T(z_1, z_2, z_3, z_4) = (9z_1, 4z_2, 0, 9z_4),$$

as you should verify. Thus the standard basis of $\mathbb{F}^4$ diagonalizes T^*T, and we see that the eigenvalues of T^*T are 9, 4, and 0. Also, the dimensions of the

Example 7.66 *continued*

eigenspaces corresponding to the eigenvalues are

$$\dim E(9, T^*T) = 2 \quad \text{and} \quad \dim E(4, T^*T) = 1 \quad \text{and} \quad \dim E(0, T^*T) = 1.$$

Taking nonnegative square roots of these eigenvalues of T^*T and using dimension information from above, we conclude that the singular values of T are $3, 3, 2, 0$.

The only eigenvalues of T are -3 and 0. 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 list of singular values does include 2.

Example 7.67 **Singular values of a linear map from $\mathbb{F}^4$ to $\mathbb{F}^3$**

Suppose $T \in \text{Hom}(\mathbb{F}^4, \mathbb{F}^3)$ has matrix (with respect to the standard bases)

$$\begin{pmatrix} 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

You can verify that the matrix of T^*T is

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 25 \end{pmatrix}$$

and that the eigenvalues of the operator T^*T are $25, 2, 0$, with $\dim E(25, T^*T) = 1$, $\dim E(2, T^*T) = 1$, and $\dim E(0, T^*T) = 2$. Thus the singular values of T are $5, \sqrt{2}, 0, 0$.

See Exercise 2 for a characterization of the positive singular values.

Result 7.68

Role of positive singular values

Suppose that $T \in \text{Hom}(V, W)$. Then

(a) T is injective $\iff$ 0 is not a singular value of T;

(b) the number of positive singular values of T equals $\dim \text{im}\, T$;

(c) T is surjective $\iff$ number of positive singular values of T equals $\dim W$.

Proof. The linear map T is injective if and only if $\ker T = \{0\}$, which happens if and only if $\ker T^*T = \{0\}$ [by 7.64(b)], which happens if and only if 0 is not an eigenvalue of T^*T, which happens if and only if 0 is not a singular value of T, completing the proof of (a).

The spectral theorem applied to T^*T shows that $\dim \text{im}\, T^*T$ equals the number of positive eigenvalues of T^*T (counting repetitions). Thus 7.64(d) implies that $\dim \text{im}\, T$ equals the number of positive singular values of T, proving (b).

Use (b) and 2.39 to show that (c) holds. □

The table below compares eigenvalues with singular values.

list of eigenvalues	list of singular values
context: vector spaces	context: inner product spaces
defined only for linear maps from a vector space to itself	defined for linear maps from an inner product space to a possibly different inner product space
can be arbitrary real numbers (if $\mathbb{F} = \mathbb{R}$) or complex numbers (if $\mathbb{F} = \mathbb{C}$)	are nonnegative numbers
can be the empty list if $\mathbb{F} = \mathbb{R}$	length of list equals dimension of domain
includes 0 $\iff$ operator is not invertible	includes 0 $\iff$ linear map is not injective
no standard order, especially if $\mathbb{F} = \mathbb{C}$	always listed in decreasing order

The next result nicely characterizes isometries in terms of singular values.

Isometries characterized by having all singular values equal 1 **Result 7.69**

Suppose that $S \in \text{Hom}(V, W)$. Then

$$S \text{ is an isometry} \iff \text{all singular values of } S \text{ equal 1.}$$

Proof. We have

$$\begin{aligned} S \text{ is an isometry} &\iff S^*S = I \\ &\iff \text{all eigenvalues of } S^*S \text{ equal 1} \\ &\iff \text{all singular values of } S \text{ equal 1,} \end{aligned}$$

where the first equivalence comes from 7.49 and the second equivalence comes from the spectral theorem (7.29 or 7.31) applied to the self-adjoint operator S^*S. □

The singular value decomposition is useful in computational linear algebra because good techniques exist for approximating eigenvalues and eigenvectors of positive operators such as T^*T, whose eigenvalues and eigenvectors lead to the singular value decomposition.

SVD for Linear Maps and for Matrices

The next result shows that every linear map from V to W has a remarkably clean description in terms of its singular values and orthonormal lists in V and W. In the next section we will see several important applications of the singular value decomposition (often called the SVD).

Singular value decomposition **Result 7.70**

Suppose $T \in \text{Hom}(V, W)$ and the positive singular values of T are $s_1, \dots, s_m$. Then there exist orthonormal lists $e_1, \dots, e_m$ in V and $f_1, \dots, f_m$ in W such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{7.71}$$

for every $v \in V$.

Proof. Let $s_1, \dots, s_n$ denote the singular values of T (thus $n = \dim V$). Because T^*T is a positive operator [see 7.64(a)], the spectral theorem implies that there exists an orthonormal basis $e_1, \dots, e_n$ of V with

$$T^*Te_k = s_k^2 e_k \tag{7.72}$$

for each $k = 1, \dots, n$.

For each $k = 1, \dots, m$, let

$$f_k = \frac{Te_k}{s_k}. \tag{7.73}$$

If $j, k \in \{1, \dots, m\}$, then

$$\langle f_j, f_k \rangle = \frac{1}{s_j s_k}\langle Te_j, Te_k \rangle = \frac{1}{s_j s_k}\langle e_j, T^*Te_k \rangle = \frac{s_k}{s_j}\langle e_j, e_k \rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k. \end{cases}$$

Thus $f_1, \dots, f_m$ is an orthonormal list in W.

If $k \in \{1, \dots, n\}$ and $k > m$, then $s_k = 0$ and hence $T^*Te_k = 0$ (by 7.72), which implies that $Te_k = 0$ [by 7.64(b)].

Suppose $v \in V$. Then

$$\begin{aligned} Tv &= T(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) \\ &= \langle v, e_1 \rangle Te_1 + \cdots + \langle v, e_m \rangle Te_m \\ &= s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m, \end{aligned}$$

where the last index in the first line switched from n to m in the second line because $Te_k = 0$ if $k > m$ (as noted in the paragraph above) and the third line follows from 7.73. The equation above is our desired result. □

Suppose $T \in \text{Hom}(V, W)$, the positive singular values of T are $s_1, \dots, s_m$, and $e_1, \dots, e_m$ and $f_1, \dots, f_m$ are as in the singular value decomposition 7.70. The orthonormal list $e_1, \dots, e_m$ can be extended to an orthonormal basis $e_1, \dots, e_{\dim V}$ of V and the orthonormal list $f_1, \dots, f_m$ can be extended to an orthonormal basis $f_1, \dots, f_{\dim W}$ of W. The formula 7.71 shows that

$$Te_k = \begin{cases} s_k f_k & \text{if } 1 \leq k \leq m, \\ 0 & \text{if } m < k \leq \dim V. \end{cases}$$

Thus the matrix of T with respect to the orthonormal bases $(e_1, \dots, e_{\dim V})$ and $(f_1, \dots, f_{\dim W})$ has the simple form

$$\text{Mat}\big(T, (e_1, \dots, e_{\dim V}), (f_1, \dots, f_{\dim W})\big)_{j,k} = \begin{cases} s_k & \text{if } 1 \leq j = k \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

If $\dim V = \dim W$ (as happens, for example, if $W = V$), then the matrix described in the paragraph above is a diagonal matrix. If we extend the definition of diagonal matrix as follows to apply to matrices that are not necessarily square, then we have proved the wonderful result that every linear map from V to W has a diagonal matrix with respect to appropriate orthonormal bases.

Diagonal matrix — Definition 7.74

An M-by-N matrix A is called a *diagonal matrix* if all entries of the matrix are 0 except possibly $A_{k,k}$ for $k = 1, \dots, \min(M, N)$.

The table below compares the spectral theorem (7.29 and 7.31) with the singular value decomposition (7.70).

spectral theorem	singular value decomposition
describes only self-adjoint operators (when $\mathbb{F} = \mathbb{R}$) or normal operators (when $\mathbb{F} = \mathbb{C}$)	describes arbitrary linear maps from an inner product space to a possibly different inner product space
produces a single orthonormal basis	produces two orthonormal lists, one for domain space and one for range space, that are not necessarily the same even when range space equals domain space
different proofs depending on whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$	same proof works regardless of whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$

The singular value decomposition gives us a new way to understand the adjoint and the inverse of a linear map. Specifically, the next result shows that given a singular value decomposition of a linear map $T \in \text{Hom}(V, W)$, we can obtain the adjoint of T simply by interchanging the roles of the e's and the f's (see 7.77). Similarly, we can obtain the pseudoinverse $T^\dagger$ (see 6.72) of T by interchanging the roles of the e's and the f's and replacing each positive singular value s_k of T with $1/s_k$ (see 7.78).

Recall that the pseudoinverse $T^\dagger$ in 7.78 below equals the inverse T^{-1} if T is invertible [see 6.73(a)].

Singular value decomposition of adjoint and pseudoinverse — Result 7.75

Suppose $T \in \text{Hom}(V, W)$ and the positive singular values of T are $s_1, \dots, s_m$. Suppose $e_1, \dots, e_m$ and $f_1, \dots, f_m$ are orthonormal lists in V and W such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{7.76}$$

for every $v \in V$. Then

$$T^*w = s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m \tag{7.77}$$

and

$$T^\dagger w = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} e_m \tag{7.78}$$

for every $w \in W$.

Proof. If $v \in V$ and $w \in W$ then

$$\begin{aligned}\langle Tv, w\rangle &= \big\langle s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m, w\big\rangle \\ &= s_1\langle v, e_1\rangle\langle f_1, w\rangle + \cdots + s_m\langle v, e_m\rangle\langle f_m, w\rangle \\ &= \big\langle v, s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m\big\rangle.\end{aligned}$$

This implies that

$$T^*w = s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m,$$

proving 7.77.

To prove 7.78, suppose $w \in W$. Let

$$v = \frac{\langle w, f_1\rangle}{s_1}e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m}e_m.$$

Apply T to both sides of the equation above, getting

$$\begin{aligned} Tv &= \frac{\langle w, f_1\rangle}{s_1}Te_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m}Te_m \\ &= \langle w, f_1\rangle f_1 + \cdots + \langle w, f_m\rangle f_m \\ &= P_{\operatorname{im} T}w, \end{aligned}$$

where the second line holds because 7.76 implies that $Te_k = s_k f_k$ if $k = 1, \ldots, m$, and the last line above holds because 7.76 implies that $f_1, \ldots, f_m$ spans $\operatorname{im} T$ and thus is an orthonormal basis of $\operatorname{im} T$ [and hence 6.61(i) applies]. The equation above, the observation that $v \in (\ker T)^\perp$ [see Exercise 8(b)], and the definition of $T^\dagger w$ (see 6.72) show that $v = T^\dagger w$, proving 7.78. □

Example 7.79 **Finding a singular value decomposition**

Define $T \in \operatorname{Hom}(\mathbb{F}^4, \mathbb{F}^3)$ by $T(x_1, x_2, x_3, x_4) = (-5x_4, 0, x_1 + x_2)$. We want to find a singular value decomposition of T. The matrix of T (with respect to the standard bases) is

$$\begin{pmatrix} 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Thus, as discussed in Example 7.67, the matrix of T^*T is

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 25 \end{pmatrix},$$

and the positive eigenvalues of T^*T are $25, 2$, with $\dim E(25, T^*T) = 1$ and $\dim E(2, T^*T) = 1$. Hence the positive singular values of T are $5, \sqrt{2}$.

Thus to find a singular value decomposition of T, we must find an orthonormal list e_1, e_2 in $\mathbb{F}^4$ and an orthonormal list f_1, f_2 in $\mathbb{F}^3$ such that

$$Tv = 5\langle v, e_1\rangle f_1 + \sqrt{2}\langle v, e_2\rangle f_2$$

for all $v \in \mathbb{F}^4$.

An orthonormal basis of $E(25, T^*T)$ is the vector $(0, 0, 0, 1)$; an orthonormal basis of $E(2, T^*T)$ is the vector $\big(1/\sqrt{2}, 1/\sqrt{2}, 0, 0\big)$. Thus, following the proof of

Example 7.79 *continued*

7.70, we take

$$e_1 = (0, 0, 0, 1) \quad \text{and} \quad e_2 = \left(1/\sqrt{2}, 1/\sqrt{2}, 0, 0\right)$$

and

$$f_1 = \frac{Te_1}{5} = (-1, 0, 0) \quad \text{and} \quad f_2 = \frac{Te_2}{\sqrt{2}} = (0, 0, 1).$$

Then, as expected, we see that e_1, e_2 is an orthonormal list in $\mathbb{F}^4$ and f_1, f_2 is an orthonormal list in $\mathbb{F}^3$ and

$$Tv = 5\langle v, e_1\rangle f_1 + \sqrt{2}\langle v, e_2\rangle f_2$$

for all $v \in \mathbb{F}^4$. Thus we have found a singular value decomposition of T.

The next result translates the singular value decomposition from the context of linear maps to the context of matrices. Specifically, the following result gives a factorization of an arbitrary matrix as the product of three nice matrices. The proof gives an explicit construction of these three matrices in terms of the singular value decomposition.

In the next result, the phrase "orthonormal columns" should be interpreted to mean that the columns are orthonormal with respect to the standard Euclidean inner product.

Matrix version of SVD Result 7.80

Suppose A is a p-by-n matrix of rank $m \geq 1$. Then there exist a p-by-m matrix B with orthonormal columns, an m-by-m diagonal matrix D with positive numbers on the diagonal, and an n-by-m matrix C with orthonormal columns such that

$$A = BDC^*.$$

Proof. Let $T: \mathbb{F}^n \to \mathbb{F}^p$ be the linear map whose matrix with respect to the standard bases equals A. Then $\dim \operatorname{im} T = m$ (by 3.78). Let

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{7.81}$$

be a singular value decomposition of T. Let

$B =$ the p-by-m matrix whose columns are $f_1, \ldots, f_m$,

$D =$ the m-by-m diagonal matrix whose diagonal entries are $s_1, \ldots, s_m$,

$C =$ the n-by-m matrix whose columns are $e_1, \ldots, e_m$.

Let $u_1, \ldots, u_m$ denote the standard basis of $\mathbb{F}^m$. If $k \in \{1, \ldots, m\}$ then

$$(AC - BD)u_k = Ae_k - B(s_k u_k) = s_k f_k - s_k f_k = 0.$$

Thus $AC = BD$.

Multiply both sides of this last equation by C^* (the conjugate transpose of C) on the right to get

$$ACC^* = BDC^*.$$

Note that the rows of C^* are the complex conjugates of $e_1, \dots, e_m$. Thus if $k \in \{1, \dots, m\}$, then the definition of matrix multiplication shows that $C^* e_k = u_k$; hence $CC^* e_k = e_k$. Thus $ACC^* v = Av$ for all $v \in \operatorname{span}(e_1, \dots, e_m)$.

If $v \in \big(\operatorname{span}(e_1, \dots, e_m)\big)^\perp$, then $Av = 0$ (as follows from 7.81) and $C^* v = 0$ (as follows from the definition of matrix multiplication). Hence $ACC^* v = Av$ for all $v \in \big(\operatorname{span}(e_1, \dots, e_m)\big)^\perp$.

Because ACC^* and A agree on $\operatorname{span}(e_1, \dots, e_m)$ and on $\big(\operatorname{span}(e_1, \dots, e_m)\big)^\perp$, we conclude that $ACC^* = A$. Thus the displayed equation above becomes

$$A = BDC^*,$$

as desired. □

Note that the matrix A in the result above has pn entries. In comparison, the matrices B, D, and C above have a total of

$$m(p + m + n)$$

entries. Thus if p and n are large numbers and the rank m is considerably less than p and n, then the number of entries that must be stored on a computer to represent A is considerably less than pn.

Exercises

7E.1 Suppose $T \in \operatorname{Hom}(V, W)$. Show that $T = 0$ if and only if all singular values of T are 0.

7E.2 Suppose $T \in \operatorname{Hom}(V, W)$ and $s > 0$. Prove that s is a singular value of T if and only if there exist nonzero vectors $v \in V$ and $w \in W$ such that

$$Tv = sw \quad \text{and} \quad T^* w = sv.$$

The vectors v, w satisfying both equations above are called a ***Schmidt pair****. Erhard Schmidt introduced the concept of singular values in 1907.*

7E.3 Give an example of $T \in \operatorname{End}(\mathbb{C}^2)$ such that 0 is the only eigenvalue of T and the singular values of T are $5, 0$.

7E.4 Suppose that $T \in \operatorname{Hom}(V, W)$, s_1 is the largest singular value of T, and s_n is the smallest singular value of T. Prove that

$$\{\|Tv\| : v \in V \text{ and } \|v\| = 1\} = [s_n, s_1].$$

7E.5 Suppose $T \in \text{End}(\mathbb{C}^2)$ is defined by $T(x, y) = (-4y, x)$. Find the singular values of T.

7E.6 Find the singular values of the differentiation operator $D \in \text{End}(\mathbb{R}[X]_{\leq 2})$ defined by $Dp = p'$, where the inner product on $\mathbb{R}[X]_{\leq 2}$ is as in Example 6.38.

7E.7 Suppose that $T \in \text{End}(V)$ is self-adjoint or that $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$ is normal. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, each included in this list as many times as the dimension of the corresponding eigenspace. Show that the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$, after these numbers have been sorted into decreasing order.

7E.8 Suppose $T \in \text{Hom}(V, W)$. Suppose $s_1 \geq s_2 \geq \cdots \geq s_m > 0$ and $e_1, \ldots, e_m$ is an orthonormal list in V and $f_1, \ldots, f_m$ is an orthonormal list in W such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for every $v \in V$.

(a) Prove that $f_1, \ldots, f_m$ is an orthonormal basis of $\text{im}\, T$.

(b) Prove that $e_1, \ldots, e_m$ is an orthonormal basis of $(\ker T)^\perp$.

(c) Prove that $s_1, \ldots, s_m$ are the positive singular values of T.

(d) Prove that if $k \in \{1, \ldots, m\}$, then e_k is an eigenvector of T^*T with corresponding eigenvalue s_k^2.

(e) Prove that

$$TT^*w = s_1^2\langle w, f_1\rangle f_1 + \cdots + s_m^2\langle w, f_m\rangle f_m$$

for all $w \in W$.

7E.9 Suppose $T \in \text{Hom}(V, W)$. Show that T and T^* have the same positive singular values.

7E.10 Suppose $T \in \text{Hom}(V, W)$ has singular values $s_1, \ldots, s_n$. Prove that if T is an invertible linear map, then T^{-1} has singular values

$$\frac{1}{s_n}, \ldots, \frac{1}{s_1}.$$

7E.11 Suppose that $T \in \text{Hom}(V, W)$ and $v_1, \ldots, v_n$ is an orthonormal basis of V. Let $s_1, \ldots, s_n$ denote the singular values of T.

(a) Prove that $\|Tv_1\|^2 + \cdots + \|Tv_n\|^2 = s_1^2 + \cdots + s_n^2$.

(b) Prove that if $W = V$ and T is a positive operator, then

$$\langle Tv_1, v_1\rangle + \cdots + \langle Tv_n, v_n\rangle = s_1 + \cdots + s_n.$$

See the comment after Exercise 5 in Section 7A.

7E.12

(a) Give an example of a finite-dimensional vector space and an operator T on it such that the singular values of T^2 do not equal the squares of the singular values of T.

(b) Suppose $T \in \text{End}(V)$ is normal. Prove that the singular values of T^2 equal the squares of the singular values of T.

7E.13 Suppose $T_1, T_2 \in \text{End}(V)$. Prove that T_1 and T_2 have the same singular values if and only if there exist unitary operators $S_1, S_2 \in \text{End}(V)$ such that $T_1 = S_1 T_2 S_2$.

7E.14 Suppose $T \in \text{Hom}(V, W)$. Let s_n denote the smallest singular value of T. Prove that $s_n \|v\| \le \|Tv\|$ for every $v \in V$.

7E.15 Suppose $T \in \text{End}(V)$ and $s_1 \ge \cdots \ge s_n$ are the singular values of T. Prove that if λ is an eigenvalue of T, then $s_1 \ge |\lambda| \ge s_n$.

7E.16 Suppose $T \in \text{Hom}(V, W)$. Prove that $(T^*)^\dagger = (T^\dagger)^*$.

Compare the result in this exercise to the analogous result for invertible linear maps [see 7.5(f)].

7E.17 Suppose $T \in \text{End}(V)$. Prove that T is self-adjoint if and only if $T^\dagger$ is self-adjoint.

Matrices unfold
Singular values gleam like stars
Order in chaos shines
—written by ChatGPT with input *haiku about SVD*

7F Consequences of Singular Value Decomposition

Norms of Linear Maps

The singular value decomposition leads to the following upper bound for $\|Tv\|$.

Upper bound for $\|Tv\|$ **Result 7.82**

Suppose $T \in \text{Hom}(V, W)$. Let s_1 be the largest singular value of T. Then

$$\|Tv\| \le s_1 \|v\|$$

for all $v \in V$.

For a lower bound on $\|Tv\|$, look at Exercise 14 in Section 7E.

Proof. Let $s_1, \dots, s_m$ denote the positive singular values of T, and let $e_1, \dots, e_m$ be an orthonormal list in V and $f_1, \dots, f_m$ be an orthonormal list in W that provide a singular value decomposition of T. Thus

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{7.83}$$

for all $v \in V$. Hence if $v \in V$ then

$$\begin{aligned}\|Tv\|^2 &= s_1^2|\langle v, e_1\rangle|^2 + \cdots + s_m^2|\langle v, e_m\rangle|^2 \\ &\le s_1^2(|\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2) \\ &\le s_1^2\|v\|^2,\end{aligned}$$

where the last inequality follows from Bessel's inequality (6.30). Taking square roots of both sides of the inequality above shows that $\|Tv\| \le s_1\|v\|$, as desired. □

Suppose $T \in \text{Hom}(V, W)$ and s_1 is the largest singular value of T. The result above shows that

$$\|Tv\| \le s_1 \text{ for all } v \in V \text{ with } \|v\| \le 1. \tag{7.84}$$

Taking $v = e_1$ in 7.83 shows that $Te_1 = s_1 f_1$. Because $\|f_1\| = 1$, this implies that $\|Te_1\| = s_1$. Thus because $\|e_1\| = 1$, the inequality in 7.84 leads to the equation

$$\max\{\|Tv\| : v \in V \text{ and } \|v\| \le 1\} = s_1. \tag{7.85}$$

The equation above is the motivation for the following definition, which defines the norm of T to be the left side of the equation above without needing to refer to singular values or the singular value decomposition.

Norm of a linear map, $\|\cdot\|$ **Definition 7.86**

Suppose $T \in \text{Hom}(V, W)$. Then the *norm* of T, denoted by $\|T\|$, is defined by

$$\|T\| = \max\{\|Tv\| : v \in V \text{ and } \|v\| \le 1\}.$$

In general, the maximum of an infinite set of nonnegative numbers need not exist. However, the discussion before 7.86 shows that the maximum in the definition of the norm of a linear map T from V to W does indeed exist (and equals the largest singular value of T).

We now have two different uses of the word *norm* and the notation $\|\cdot\|$. Our first use of this notation was in connection with an inner product on V, when we defined $\|v\| = \sqrt{\langle v, v\rangle}$ for each $v \in V$. Our second use of the norm notation and terminology is with the definition we just made of $\|T\|$ for $T \in \text{Hom}(V, W)$. The norm $\|T\|$ for $T \in \text{Hom}(V, W)$ does not usually come from taking an inner product of T with itself (see Exercise 21). You should be able to tell from the context and from the symbols used which meaning of the norm is intended.

The properties of the norm on $\text{Hom}(V, W)$ listed below look identical to properties of the norm on an inner product space (see 6.9 and 6.18). The inequality in (d) is called the *triangle inequality*, thus using the same terminology that we used for the norm on V. For the reverse triangle inequality, see Exercise 1.

Basic properties of norms of linear maps **Result 7.87**

Suppose $T \in \text{Hom}(V, W)$. Then

(a) $\|T\| \geq 0$;

(b) $\|T\| = 0 \iff T = 0$;

(c) $\|\lambda T\| = |\lambda| \, \|T\|$ for all $\lambda \in \mathbb{F}$;

(d) $\|S + T\| \leq \|S\| + \|T\|$ for all $S \in \text{Hom}(V, W)$.

Proof.

(a) Because $\|Tv\| \geq 0$ for every $v \in V$, the definition of $\|T\|$ implies that $\|T\| \geq 0$.

(b) Suppose $\|T\| = 0$. Thus $Tv = 0$ for all $v \in V$ with $\|v\| \leq 1$. If $u \in V$ with $u \neq 0$, then

$$Tu = \|u\| \, T\Big(\frac{u}{\|u\|}\Big) = 0,$$

where the last equality holds because $u/\|u\|$ has norm 1. Because $Tu = 0$ for all $u \in V$, we have $T = 0$.

Conversely, if $T = 0$ then $Tv = 0$ for all $v \in V$ and hence $\|T\| = 0$.

(c) Suppose $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} \|\lambda T\| &= \max\{\|\lambda Tv\| : v \in V \text{ and } \|v\| \leq 1\} \\ &= |\lambda| \max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\} \\ &= |\lambda| \, \|T\|. \end{aligned}$$

(d) Suppose $S \in \text{Hom}(V, W)$. The definition of $\|S + T\|$ implies that there exists $v \in V$ such that $\|v\| \leq 1$ and $\|S + T\| = \|(S + T)v\|$. Now

$$\|S + T\| = \|(S + T)v\| = \|Sv + Tv\| \leq \|Sv\| + \|Tv\| \leq \|S\| + \|T\|,$$

completing the proof of (d). □

For $S, T \in \text{Hom}(V, W)$, the quantity $\|S - T\|$ is often called the distance between S and T. Informally, think of the condition that $\|S - T\|$ is a small number as meaning that S and T are close together. For example, Exercise 9 asserts that for every $T \in \text{End}(V)$, there is an invertible operator as close to T as we wish.

Alternative formulas for $\|T\|$ **Result 7.88**

Suppose $T \in \text{Hom}(V, W)$. Then

(a) $\|T\|$ = the largest singular value of T;

(b) $\|T\| = \max\{\|Tv\| : v \in V \text{ and } \|v\| = 1\}$;

(c) $\|T\|$ = the smallest number c such that $\|Tv\| \leq c\|v\|$ for all $v \in V$.

Proof.

(a) See 7.85.

(b) Let $v \in V$ be such that $0 < \|v\| \le 1$. Let $u = v/\|v\|$. Then

$$\|u\| = \left\| \frac{v}{\|v\|} \right\| = 1 \quad \text{and} \quad \|Tu\| = \left\| T\Big(\frac{v}{\|v\|}\Big) \right\| = \frac{\|Tv\|}{\|v\|} \ge \|Tv\|.$$

Thus when finding the maximum of $\|Tv\|$ with $\|v\| \le 1$, we can restrict attention to vectors in V with norm 1, proving (b).

(c) Suppose $v \in V$ and $v \ne 0$. Then the definition of $\|T\|$ implies that

$$\left\| T\Big(\frac{v}{\|v\|}\Big) \right\| \le \|T\|,$$

which implies that

$$\|Tv\| \le \|T\| \, \|v\|. \tag{7.89}$$

Now suppose $c \ge 0$ and $\|Tv\| \le c\|v\|$ for all $v \in V$. This implies that

$$\|Tv\| \le c$$

for all $v \in V$ with $\|v\| \le 1$. Taking the maximum of the left side of the inequality above over all $v \in V$ with $\|v\| \le 1$ shows that $\|T\| \le c$. Thus $\|T\|$ is the smallest number c such that $\|Tv\| \le c\|v\|$ for all $v \in V$. □

When working with norms of linear maps, you will probably frequently use the inequality 7.89.

For computing an approximation of the norm of a linear map T given the matrix of T with respect to some orthonormal bases, 7.88(a) is likely to be most useful. The matrix of T^*T is quickly computable from matrix multiplication. Then a computer can be asked to find an approximation for the largest eigenvalue of T^*T (excellent numerical algorithms exist for this purpose). Then taking the square root and using 7.88(a) gives an approximation for the norm of T (which usually cannot be computed exactly).

You should verify all assertions in the example below.

Example 7.90 **Norms**

- If I denotes the usual identity operator on V, then $\|I\| = 1$.
- If $T \in \text{End}(\mathbb{F}^n)$ and the matrix of T with respect to the standard basis of $\mathbb{F}^n$ consists of all 1's, then $\|T\| = n$.
- If $T \in \text{End}(V)$ and V has an orthonormal basis consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, then $\|T\|$ is the maximum of the numbers $|\lambda_1|, \dots, |\lambda_n|$.
- Suppose $T \in \text{End}(\mathbb{R}^5)$ is the operator whose matrix (with respect to the standard basis) is the 5-by-5 matrix whose entry in row j, column k is $1/(j^2 + k)$. Standard mathematical software shows that the largest singular value of T is approximately 0.8 and the smallest singular value of T is approximately 10^{-6}.

Example 7.90 *continued*

Thus $\|T\| \approx 0.8$ and (using Exercise 10 in Section 7E) $\|T^{-1}\| \approx 10^6$. It is not possible to find exact formulas for these norms.

A linear map and its adjoint have the same norm, as shown by the next result.

Norm of the adjoint Result 7.91

Suppose $T \in \text{Hom}(V, W)$. Then $\|T^*\| = \|T\|$.

Proof. Suppose $w \in W$. Then

$$\|T^*w\|^2 = \langle T^*w, T^*w\rangle = \langle TT^*w, w\rangle \le \|TT^*w\|\,\|w\| \le \|T\|\,\|T^*w\|\,\|w\|.$$

The inequality above implies that

$$\|T^*w\| \le \|T\|\,\|w\|,$$

which along with 7.88(c) implies that $\|T^*\| \le \|T\|$.

Replacing T with T^* in the inequality $\|T^*\| \le \|T\|$ and then using the equation $(T^*)^* = T$ shows that $\|T\| \le \|T^*\|$. Thus $\|T^*\| = \|T\|$, as desired. □

You may want to construct an alternative proof of the result above using Exercise 9 in Section 7E, which asserts that a linear map and its adjoint have the same positive singular values.

Approximation by Linear Maps with Lower-Dimensional Range

The next result is a spectacular application of the singular value decomposition. It says that to best approximate a linear map by a linear map whose range has dimension at most k, chop off the singular value decomposition after the first k terms. Specifically, the linear map T_k in the next result has the property that $\dim \text{im}\, T_k = k$ and T_k minimizes the distance to T among all linear maps with range of dimension at most k. This result leads to algorithms for compressing huge matrices while preserving their most important information.

Best approximation by linear map whose range has dimension $\le k$ Result 7.92

Suppose $T \in \text{Hom}(V, W)$ and $s_1 \ge \cdots \ge s_m$ are the positive singular values of T. Suppose $1 \le k < m$. Then

$$\min\{\|T - S\| : S \in \text{Hom}(V, W) \text{ and } \dim \text{im}\, S \le k\} = s_{k+1}.$$

Furthermore, if

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

is a singular value decomposition of T and $T_k \in \text{Hom}(V, W)$ is defined by

$$T_k v = s_1\langle v, e_1\rangle f_1 + \cdots + s_k\langle v, e_k\rangle f_k$$

for each $v \in V$, then $\dim \text{im}\, T_k = k$ and $\|T - T_k\| = s_{k+1}$.

Proof. If $v \in V$ then

$$\begin{aligned}\|(T - T_k)v\|^2 &= \|s_{k+1}\langle v, e_{k+1}\rangle f_{k+1} + \cdots + s_m\langle v, e_m\rangle f_m\|^2 \\ &= s_{k+1}^2|\langle v, e_{k+1}\rangle|^2 + \cdots + s_m^2|\langle v, e_m\rangle|^2 \\ &\le s_{k+1}^2(|\langle v, e_{k+1}\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2) \\ &\le s_{k+1}^2\|v\|^2.\end{aligned}$$

Thus $\|T - T_k\| \le s_{k+1}$. The equation $(T - T_k)e_{k+1} = s_{k+1}f_{k+1}$ now shows that $\|T - T_k\| = s_{k+1}$.

Suppose $S \in \text{Hom}(V, W)$ and $\dim \text{im}\, S \le k$. Thus $Se_1, \ldots, Se_{k+1}$, which is a list of length $k + 1$, is linearly dependent. Hence there exist $a_1, \ldots, a_{k+1} \in \mathbb{F}$, not all 0, such that

$$a_1Se_1 + \cdots + a_{k+1}Se_{k+1} = 0.$$

Now $a_1e_1 + \cdots + a_{k+1}e_{k+1} \ne 0$ because $a_1, \ldots, a_{k+1}$ are not all 0. We have

$$\begin{aligned}\|(T - S)(a_1e_1 + \cdots + a_{k+1}e_{k+1})\|^2 &= \|T(a_1e_1 + \cdots + a_{k+1}e_{k+1})\|^2 \\ &= \|s_1a_1f_1 + \cdots + s_{k+1}a_{k+1}f_{k+1}\|^2 \\ &= s_1^2|a_1|^2 + \cdots + s_{k+1}^2|a_{k+1}|^2 \\ &\ge s_{k+1}^2(|a_1|^2 + \cdots + |a_{k+1}|^2) \\ &= s_{k+1}^2\|a_1e_1 + \cdots + a_{k+1}e_{k+1}\|^2.\end{aligned}$$

Because $a_1e_1 + \cdots + a_{k+1}e_{k+1} \ne 0$, the inequality above implies that

$$\|T - S\| \ge s_{k+1}.$$

Thus $S = T_k$ minimizes $\|T - S\|$ among $S \in \text{Hom}(V, W)$ with $\dim \text{im}\, S \le k$. □

For other examples of the use of the singular value decomposition in best approximation, see Exercise 22, which finds a subspace of given dimension on which the restriction of a linear map is as small as possible, and Exercise 27, which finds a unitary operator that is as close as possible to a given operator.

Polar Decomposition

Recall our discussion before 7.54 of the analogy between complex numbers z with $|z| = 1$ and unitary operators. Continuing with this analogy, note that every complex number z except 0 can be written in the form

$$\begin{aligned}z &= \Big(\frac{z}{|z|}\Big)|z| \\ &= \Big(\frac{z}{|z|}\Big)\sqrt{\overline{z}z},\end{aligned}$$

where the first factor, namely, $z/|z|$, has absolute value 1.

Our analogy leads us to guess that every operator $T \in \text{End}(V)$ can be written as a unitary operator times $\sqrt{T^*T}$. That guess is indeed correct. The corresponding result

is called the polar decomposition, which gives a beautiful description of an arbitrary operator on V.

Note that if $T \in \text{End}(V)$, then T^*T is a positive operator [as was shown in 7.64(a)]. Thus the operator $\sqrt{T^*T}$ makes sense and is well defined as a positive operator on V.

The polar decomposition that we are about to state and prove says that every operator on V is the product of a unitary operator and a positive operator. Thus we can write an arbitrary operator on V as the product of two nice operators, each of which comes from a class that we can completely describe and that we understand reasonably well. The unitary operators are described by 7.55 if $\mathbb{F} = \mathbb{C}$; the positive operators are described by the real and complex spectral theorems (7.29 and 7.31).

Specifically, consider the case $\mathbb{F} = \mathbb{C}$, and suppose

$$T = S\sqrt{T^*T}$$

is a polar decomposition of an operator $T \in \text{End}(V)$, where S is a unitary operator. 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—S may require one orthonormal basis and $\sqrt{T^*T}$ may require a different orthonormal basis.

However (still assuming that $\mathbb{F} = \mathbb{C}$), if T is normal, then an orthonormal basis of V can be chosen such that both S and $\sqrt{T^*T}$ have diagonal matrices with respect to this basis—see Exercise 31. The converse is also true: If $T \in \text{End}(V)$ and $T = S\sqrt{T^*T}$ for some unitary operator $S \in \text{End}(V)$ such that S and $\sqrt{T^*T}$ both have diagonal matrices with respect to the same orthonormal basis of V, then T is normal. This holds because T then has a diagonal matrix with respect to this same orthonormal basis, which implies that T is normal [by the equivalence of (c) and (a) in 7.31].

The polar decomposition below is valid on both real and complex inner product spaces and for all operators on those spaces.

Polar decomposition **Result 7.93**

Suppose $T \in \text{End}(V)$. Then there exists a unitary operator $S \in \text{End}(V)$ such that

$$T = S\sqrt{T^*T}.$$

Proof. Let $s_1, \dots, s_m$ be the positive singular values of T, and let $e_1, \dots, e_m$ and $f_1, \dots, f_m$ be orthonormal lists in V such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{7.94}$$

for every $v \in V$. Extend $e_1, \dots, e_m$ and $f_1, \dots, f_m$ to orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ of V.

Define $S \in \text{End}(V)$ by

$$Sv = \langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n$$

for each $v \in V$. Then

$$\begin{aligned}\|Sv\|^2 &= \|\langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n\|^2 \\ &= |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 \\ &= \|v\|^2.\end{aligned}$$

Thus S is a unitary operator.

Applying T^* to both sides of 7.94 and then using the formula for T^* given by 7.77 shows that

$$T^*Tv = s_1^2\langle v, e_1\rangle e_1 + \cdots + s_m^2\langle v, e_m\rangle e_m$$

for every $v \in V$. Thus if $v \in V$, then

$$\sqrt{T^*T}\, v = s_1\langle v, e_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m$$

because the operator that sends v to the right side of the equation above is a positive operator whose square equals T^*T. Now

$$\begin{aligned}S\sqrt{T^*T}\, v &= S(s_1\langle v, e_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m) \\ &= s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \\ &= Tv,\end{aligned}$$

where the last equation follows from 7.94. □

Exercise 27 shows that the unitary operator S produced in the proof above is as close as a unitary operator can be to T.

Alternative proofs of the polar decomposition directly use the spectral theorem, avoiding the singular value decomposition. However, the proof above seems cleaner than those alternative proofs.

Operators Applied to Ellipsoids and Parallelepipeds

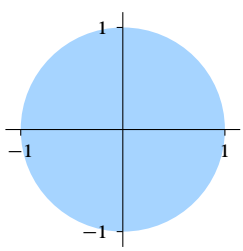
Ball, B **Definition 7.95**

The *ball* in V of radius 1 centered at 0, denoted by B, is defined by

$$B = \{v \in V : \|v\| < 1\}.$$

If $\dim V = 2$, the word *disk* is sometimes used instead of *ball*. However, using *ball* in all dimensions is less confusing. Similarly, if $\dim V = 2$, then the word *ellipse* is sometimes used instead of the word *ellipsoid* that we are about to define. Again, using *ellipsoid* in all dimensions is less confusing.

You can think of the ellipsoid defined below as obtained by starting with the ball B and then stretching by a factor of s_k along each f_k-axis.



The ball B in $\mathbb{R}^2$.

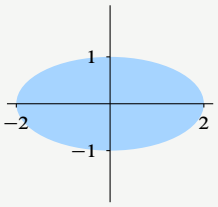
Definition 7.96 **Ellipsoid, $E(s_1 f_1, \dots, s_n f_n)$, principal axes**

Suppose that $f_1, \dots, f_n$ is an orthonormal basis of V and $s_1, \dots, s_n$ are positive numbers. The *ellipsoid* $E(s_1 f_1, \dots, s_n f_n)$ with *principal axes* $s_1 f_1, \dots, s_n f_n$ is defined by

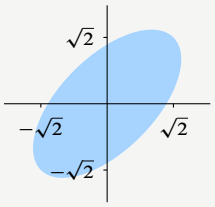
$$E(s_1 f_1, \dots, s_n f_n) = \left\{ v \in V : \frac{|\langle v, f_1 \rangle|^2}{s_1^2} + \cdots + \frac{|\langle v, f_n \rangle|^2}{s_n^2} < 1 \right\}.$$

The ellipsoid notation $E(s_1 f_1, \dots, s_n f_n)$ does not explicitly include the inner product space V, even though the definition above depends on V. However, the inner product space V should be clear from the context and also from the requirement that $f_1, \dots, f_n$ be an orthonormal basis of V.

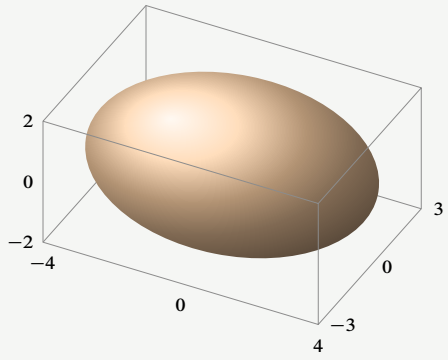
Example 7.97 **Ellipsoids**



The ellipsoid $E(2f_1, f_2)$ in $\mathbb{R}^2$, where f_1, f_2 is the standard basis of $\mathbb{R}^2$.



The ellipsoid $E(2f_1, f_2)$ in $\mathbb{R}^2$, where $f_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and $f_2 = (-1/\sqrt{2}, 1/\sqrt{2})$.



The ellipsoid $E(4f_1, 3f_2, 2f_3)$ in $\mathbb{R}^3$, where f_1, f_2, f_3 is the standard basis of $\mathbb{R}^3$.

The ellipsoid $E(f_1, \dots, f_n)$ equals the ball B in V for every orthonormal basis $f_1, \dots, f_n$ of V [by Parseval's identity 6.34(b)].

Notation 7.98

$T(\Omega)$

For T a function defined on V and $\Omega \subseteq V$, define $T(\Omega)$ by

$$T(\Omega) = \{Tv : v \in \Omega\}.$$

Thus if T is a function defined on V, then $T(V) = \operatorname{im} T$.

The next result states that every invertible operator $T \in \operatorname{End}(V)$ maps the ball B in V onto an ellipsoid in V. The proof shows that the principal axes of this ellipsoid come from the singular value decomposition of T.

Invertible operator takes ball to ellipsoid **Result 7.99**

Suppose $T \in \operatorname{End}(V)$ is invertible. Then T maps the ball B in V onto an ellipsoid in V.

Proof. Suppose T has singular value decomposition

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n \tag{7.100}$$

for all $v \in V$; here $s_1, \ldots, s_n$ are the singular values of T and $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ are both orthonormal bases of V. We will show that $T(B) = E(s_1 f_1, \ldots, s_n f_n)$.

First suppose $v \in B$. Because T is invertible, none of the singular values $s_1, \ldots, s_n$ equals 0 (see 7.68). Thus 7.100 implies that

$$\frac{|\langle Tv, f_1\rangle|^2}{s_1^2} + \cdots + \frac{|\langle Tv, f_n\rangle|^2}{s_n^2} = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 < 1.$$

Thus $Tv \in E(s_1 f_1, \ldots, s_n f_n)$. Hence $T(B) \subseteq E(s_1 f_1, \ldots, s_n f_n)$.

To prove inclusion in the other direction, now suppose $w \in E(s_1 f_1, \ldots, s_n f_n)$. Let

$$v = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_n\rangle}{s_n} e_n.$$

Then $\|v\| < 1$ and 7.100 implies that $Tv = \langle w, f_1\rangle f_1 + \cdots + \langle w, f_n\rangle f_n = w$. Thus $T(B) \supseteq E(s_1 f_1, \ldots, s_n f_n)$. □

We now use the previous result to show that invertible operators take all ellipsoids, not just the ball of radius 1, to ellipsoids.

Invertible operator takes ellipsoids to ellipsoids **Result 7.101**

Suppose $T \in \operatorname{End}(V)$ is invertible and E is an ellipsoid in V. Then $T(E)$ is an ellipsoid in V.

Proof. There exist an orthonormal basis $f_1, \ldots, f_n$ of V and positive numbers $s_1, \ldots, s_n$ such that $E = E(s_1 f_1, \ldots, s_n f_n)$. Define $S \in \operatorname{End}(V)$ by

$$S(a_1 f_1 + \cdots + a_n f_n) = a_1 s_1 f_1 + \cdots + a_n s_n f_n.$$

Then S maps the ball B of V onto E, as you can verify. Thus

$$T(E) = T\big(S(B)\big) = (TS)(B).$$

The equation above and 7.99, applied to TS, show that $T(E)$ is an ellipsoid in V. □

Recall (see 3.95) that if $u \in V$ and $\Omega \subseteq V$ then $u + \Omega$ is defined by

$$u + \Omega = \{u + w : w \in \Omega\}.$$

Geometrically, the sets Ω and $u + \Omega$ look the same, but they are in different locations.

In the following definition, if $\dim V = 2$ then the word *parallelogram* is often used instead of *parallelepiped*.

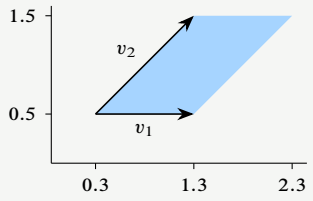
$P(v_1, \dots, v_n)$, parallelepiped **Definition 7.102**

Suppose $v_1, \dots, v_n$ is a basis of V. Let

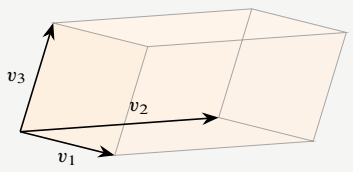
$$P(v_1, \dots, v_n) = \{a_1 v_1 + \cdots + a_n v_n : a_1, \dots, a_n \in (0, 1)\}.$$

A *parallelepiped* is a set of the form $u + P(v_1, \dots, v_n)$ for some $u \in V$. The vectors $v_1, \dots, v_n$ are called the *edges* of this parallelepiped.

Example 7.103 **Parallelepipeds**



The parallelepiped $(0.3, 0.5) + P\big((1, 0), (1, 1)\big)$ in $\mathbb{R}^2$.



A parallelepiped in $\mathbb{R}^3$.

Invertible operator takes parallelepipeds to parallelepipeds **Result 7.104**

Suppose $u \in V$, $v_1, \dots, v_n$ is a basis of V, and $T \in \operatorname{End}(V)$ is invertible. Then

$$T\big(u + P(v_1, \dots, v_n)\big) = Tu + P(Tv_1, \dots, Tv_n).$$

Proof. Because T is invertible, the list $Tv_1, \dots, Tv_n$ is a basis of V. The linearity of T implies that

$$T(u + a_1 v_1 + \cdots + a_n v_n) = Tu + a_1 Tv_1 + \cdots + a_n Tv_n$$

for all $a_1, \dots, a_n \in (0, 1)$. Thus $T\big(u + P(v_1, \dots, v_n)\big) = Tu + P(Tv_1, \dots, Tv_n)$. □

Just as the rectangles are distinguished among the parallelograms in $\mathbb{R}^2$, we give a special name to the parallelepipeds in V whose defining edges are orthogonal to each other.

Box — Definition 7.105

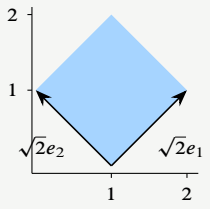
A *box* in V is a set of the form

$$u + P(r_1 e_1, \dots, r_n e_n),$$

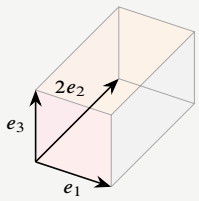
where $u \in V$ and $r_1, \dots, r_n$ are positive numbers and $e_1, \dots, e_n$ is an orthonormal basis of V.

Note that in the special case of $\mathbb{R}^2$ each box is a rectangle, but the terminology *box* can be used in all dimensions.

Example 7.106 Boxes



The box $(1, 0) + P(\sqrt{2}e_1, \sqrt{2}e_2)$, where $e_1 = (1/\sqrt{2}, 1/\sqrt{2})$ and $e_2 = (-1/\sqrt{2}, 1/\sqrt{2})$.



The box $P(e_1, 2e_2, e_3)$, where e_1, e_2, e_3 is the standard basis of $\mathbb{R}^3$.

Suppose $T \in \text{End}(V)$ is invertible. Then T maps every parallelepiped in V to a parallelepiped in V (by 7.104). In particular, T maps every box in V to a parallelepiped in V. This raises the question of whether T maps some boxes in V to boxes in V. The following result answers this question, with the help of the singular value decomposition.

Every invertible operator takes some boxes to boxes — Result 7.107

Suppose $T \in \text{End}(V)$ is invertible. Suppose T has singular value decomposition

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n,$$

where $s_1, \dots, s_n$ are the singular values of T and $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are orthonormal bases of V and the equation above holds for all $v \in V$. Then T maps the box $u + P(r_1 e_1, \dots, r_n e_n)$ onto the box $Tu + P(r_1 s_1 f_1, \dots, r_n s_n f_n)$ for all positive numbers $r_1, \dots, r_n$ and all $u \in V$.

Proof. If $a_1, \dots, a_n \in (0, 1)$ and $r_1, \dots, r_n$ are positive numbers and $u \in V$, then

$$T(u + a_1 r_1 e_1 + \cdots + a_n r_n e_n) = Tu + a_1 r_1 s_1 f_1 + \cdots + a_n r_n s_n f_n.$$

Thus $T\big(u + P(r_1 e_1, \dots, r_n e_n)\big) = Tu + P(r_1 s_1 f_1, \dots, r_n s_n f_n)$. □

Volume via Singular Values

Our goal in this subsection is to understand how an operator changes the volume of subsets of its domain. Because notions of volume belong to analysis rather than to linear algebra, we will work only with an intuitive notion of volume. Our intuitive approach to volume can be converted into appropriate correct definitions, correct statements, and correct proofs using the machinery of analysis.

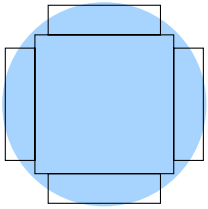
Our intuition about volume works best in real inner product spaces. Thus the assumption that $\mathbb{F} = \mathbb{R}$ will appear frequently in the rest of this subsection.

If $\dim V = n$, then by *volume* we will mean n-dimensional volume. You should be familiar with this concept in $\mathbb{R}^3$. When $n = 2$, this is usually called area instead of volume, but for consistency we use the word volume in all dimensions. The most fundamental intuition about volume is that the volume of a box (whose defining edges are by definition orthogonal to each other) is the product of the lengths of the defining edges. Thus we make the following definition.

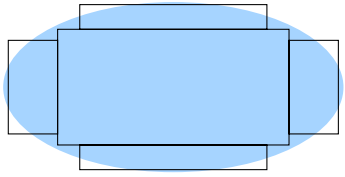
Volume of a box — Definition 7.108

Suppose $\mathbb{F} = \mathbb{R}$. If $u \in V$ and $r_1, \dots, r_n$ are positive numbers and $e_1, \dots, e_n$ is an orthonormal basis of V, then

$$\text{volume}\big(u + P(r_1 e_1, \dots, r_n e_n)\big) = r_1 \times \cdots \times r_n.$$



Volume of this ball ≈ sum of the volumes of the five boxes.



Each box here has twice the width and the same height as the boxes in the previous figure.

The definition above agrees with the familiar formulas for the area (which we are calling the volume) of a rectangle in $\mathbb{R}^2$ and for the volume of a box in $\mathbb{R}^3$. For example, the first box in Example 7.106 has two-dimensional volume (or area) 2 because the defining edges of that box have length $\sqrt{2}$ and $\sqrt{2}$. The second box in Example 7.106 has three-dimensional volume 2 because the defining edges of that box have length 1, 2, and 1.

To define the volume of a subset of V, approximate the subset by a finite collection of disjoint boxes, and then add up the volumes of the approximating collection of boxes. As we approximate a subset of V more accurately by disjoint unions of more boxes, we get a better approximation to the volume.

These ideas should remind you of how the Riemann integral is defined by approximating the area under a curve by a disjoint collection of rectangles. This discussion leads to the following nonrigorous but intuitive definition.

Volume — Definition 7.109

Suppose $\mathbb{F} = \mathbb{R}$ and $\Omega \subseteq V$. Then the *volume* of Ω, denoted by $\text{volume}\,\Omega$, is approximately the sum of the volumes of a collection of disjoint boxes that approximate Ω.

We are ignoring many reasonable questions by taking an intuitive approach to volume. For example, if we approximate Ω by boxes with respect to one basis, do we get the same volume if we approximate Ω by boxes with respect to a different basis? If Ω_1 and Ω_2 are disjoint subsets of V, is $\text{volume}(\Omega_1 \cup \Omega_2) = \text{volume}\,\Omega_1 + \text{volume}\,\Omega_2$?

Provided that we consider only reasonably nice subsets of V, techniques of analysis show that both these questions have affirmative answers that agree with our intuition about volume.

Example 7.110 **Volume change by a linear map**

Suppose that $T \in \text{End}(\mathbb{R}^2)$ is defined by $Tv = 2\langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2$, where e_1, e_2 is the standard basis of $\mathbb{R}^2$. This linear map stretches vectors along the e_1-axis by a factor of 2 and leaves vectors along the e_2-axis unchanged. The ball approximated by five boxes above gets mapped by T to the ellipsoid shown here. Each of the five boxes in the original figure gets mapped to a box of twice the width and the same height as in the original figure. Hence each box gets mapped to a box of twice the volume (area) as in the original figure. The sum of the volumes of the five new boxes approximates the volume of the ellipsoid. Thus T changes the volume of the ball by a factor of 2.

In the example above, T maps boxes with respect to the basis e_1, e_2 to boxes with respect to the same basis; thus we can see how T changes volume. In general, an operator maps boxes to parallelepipeds that are not boxes. However, if we choose the right basis (coming from the singular value decomposition!), then boxes with respect to that basis get mapped to boxes with respect to a possibly different basis, as shown in 7.107. This observation leads to a natural proof of the following result.

Volume changes by a factor of the product of the singular values **Result 7.111**
Suppose $\mathbb{F} = \mathbb{R}$, $T \in \text{End}(V)$ is invertible, and $\Omega \subseteq V$. Then

$$\text{volume } T(\Omega) = (\text{product of singular values of } T)(\text{volume } \Omega).$$

Proof. Suppose T has singular value decomposition

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for all $v \in V$, where $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ are orthonormal bases of V.

Approximate Ω by boxes of the form $u + P(r_1 e_1, \ldots, r_n e_n)$, which have volume $r_1 \times \cdots \times r_n$. The operator T maps each box $u + P(r_1 e_1, \ldots, r_n e_n)$ onto the box $Tu + P(r_1 s_1 f_1, \ldots, r_n s_n f_n)$, which has volume $(s_1 \times \cdots \times s_n)(r_1 \times \cdots \times r_n)$.

The operator T maps a collection of boxes that approximate Ω onto a collection of boxes that approximate $T(\Omega)$. Because T changes the volume of each box in a collection that approximates Ω by a factor of $s_1 \times \cdots \times s_n$, the linear map T changes the volume of Ω by the same factor. □

Suppose $T \in \text{End}(V)$. As we will see when we get to determinants, the product of the singular values of T equals $|\det T|$; see 9.60 and 9.61.

Properties of an Operator as Determined by Its Eigenvalues

We conclude this chapter by presenting the table below. The context of this table is a finite-dimensional complex inner product space. The first column of the table shows a property that a normal operator on such a space might have. The second column of the table shows a subset of $\mathbb{C}$ such that the operator has the corresponding property if and only if all eigenvalues of the operator lie in the specified subset. For example, the first row of the table states that a normal operator is invertible if and only if all its eigenvalues are nonzero (this first row is the only one in the table that does not need the hypothesis that the operator is normal).

Make sure you can explain why all results in the table hold. For example, the last row of the table holds because the norm of an operator equals its largest singular value (by 7.85) and the singular values of a normal operator, assuming $\mathbb{F} = \mathbb{C}$, equal the absolute values of the eigenvalues (by Exercise 7 in Section 7E).

properties of a normal operator	**eigenvalues are contained in**
invertible	$\mathbb{C} \setminus \{0\}$
self-adjoint	$\mathbb{R}$
skew	$\{\lambda \in \mathbb{C} : \operatorname{Re}\lambda = 0\}$
orthogonal projection	$\{0, 1\}$
positive	$[0, \infty)$
unitary	$\{\lambda \in \mathbb{C} : |\lambda| = 1\}$
norm is less than 1	$\{\lambda \in \mathbb{C} : |\lambda| < 1\}$

Exercises

7F.1 Prove that if $S, T \in \operatorname{Hom}(V, W)$, then $\big|\|S\| - \|T\|\big| \le \|S - T\|$.

The inequality above is called the ***reverse triangle inequality****.*

7F.2 Suppose that $T \in \operatorname{End}(V)$ is self-adjoint or that $\mathbb{F} = \mathbb{C}$ and $T \in \operatorname{End}(V)$ is normal. Prove that

$$\|T\| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

7F.3 Suppose $T \in \operatorname{Hom}(V, W)$ and $v \in V$. Prove that

$$\|Tv\| = \|T\|\,\|v\| \iff T^*Tv = \|T\|^2 v.$$

7F.4 Suppose $T \in \operatorname{Hom}(V, W)$, $v \in V$, and $\|Tv\| = \|T\|\,\|v\|$. Prove that if $u \in V$ and $\langle u, v\rangle = 0$, then $\langle Tu, Tv\rangle = 0$.

7F.5 Suppose U is a finite-dimensional inner product space, $T \in \text{Hom}(V, U)$, and $S \in \text{Hom}(U, W)$. Prove that

$$\|ST\| \le \|S\| \, \|T\|.$$

7F.6 Prove or give a counterexample: If $S, T \in \text{End}(V)$, then $\|ST\| = \|TS\|$.

7F.7 Show that defining $d(S, T) = \|S - T\|$ for $S, T \in \text{Hom}(V, W)$ makes d a metric on $\text{Hom}(V, W)$.

This exercise is intended for readers who are familiar with metric spaces.

7F.8

(a) Prove that if $T \in \text{End}(V)$ and $\|I - T\| < 1$, then T is invertible.

(b) Suppose that $S \in \text{End}(V)$ is invertible. Prove that if $T \in \text{End}(V)$ and $\|S - T\| < 1/\|S^{-1}\|$, then T is invertible.

This exercise shows that the set of invertible operators in $\text{End}(V)$ *is an open subset of* $\text{End}(V)$, *using the metric defined in Exercise 7.*

7F.9 Suppose $T \in \text{End}(V)$. Prove that for every $\epsilon > 0$, there exists an invertible operator $S \in \text{End}(V)$ such that $0 < \|T - S\| < \epsilon$.

7F.10 Suppose $\dim V > 1$ and $T \in \text{End}(V)$ is not invertible. Prove that for every $\epsilon > 0$, there exists $S \in \text{End}(V)$ such that $0 < \|T - S\| < \epsilon$ and S is not invertible.

7F.11 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that for every $\epsilon > 0$ there exists a diagonalizable operator $S \in \text{End}(V)$ such that $0 < \|T - S\| < \epsilon$.

7F.12 Suppose $T \in \text{End}(V)$ is a positive operator. Show that $\left\|\sqrt{T}\right\| = \sqrt{\|T\|}$.

7F.13 Suppose $S, T \in \text{End}(V)$ are positive operators. Show that

$$\|S - T\| \le \max\{\|S\|, \|T\|\} \le \|S + T\|.$$

7F.14 Suppose U and W are subspaces of V such that $\|P_U - P_W\| < 1$. Prove that $\dim U = \dim W$.

7F.15 Define $T \in \text{End}(\mathbb{F}^3)$ by

$$T(z_1, z_2, z_3) = (z_3, 2z_1, 3z_2).$$

Find (explicitly) a unitary operator $S \in \text{End}(\mathbb{F}^3)$ such that $T = S\sqrt{T^*T}$.

7F.16 Suppose $S \in \text{End}(V)$ is a positive invertible operator. Prove that there exists $\delta > 0$ such that T is a positive operator for every self-adjoint operator $T \in \text{End}(V)$ with $\|S - T\| < \delta$.

7F.17 Prove that if $u \in V$ and φ_u is the linear functional on V defined by the equation $\varphi_u(v) = \langle v, u\rangle$, then $\|\varphi_u\| = \|u\|$.

Here we are thinking of the scalar field $\mathbb{F}$ as an inner product space with $\langle \alpha, \beta\rangle = \alpha\overline{\beta}$ for all $\alpha, \beta \in \mathbb{F}$. Thus $\|\varphi_u\|$ means the norm of φ_u as a linear map from V to $\mathbb{F}$.

7F.18 Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $T \in \text{Hom}(V, W)$.

(a) Prove that $\max\{\|Te_1\|, \dots, \|Te_n\|\} \le \|T\| \le (\|Te_1\|^2 + \cdots + \|Te_n\|^2)^{1/2}$.

(b) Prove that $\|T\| = (\|Te_1\|^2 + \cdots + \|Te_n\|^2)^{1/2}$ if and only if $\dim \text{im}\, T \le 1$.

Here $e_1, \dots, e_n$ is an arbitrary orthonormal basis of V, not necessarily connected with a singular value decomposition of T. If $s_1, \dots, s_n$ is the list of singular values of T, then the right side of the inequality above equals $(s_1^2 + \cdots + s_n^2)^{1/2}$, as was shown in Exercise 11(a) in Section 7E.

7F.19 Prove that if $T \in \text{Hom}(V, W)$, then $\|T^*T\| = \|T\|^2$.

This formula for $\|T^*T\|$ leads to the important subject of C^*-algebras.

7F.20 Suppose $T \in \text{End}(V)$ is normal. Prove that $\|T^k\| = \|T\|^k$ for every positive integer k.

7F.21 Suppose $\dim V > 1$ and $\dim W > 1$. Prove that the norm on $\text{Hom}(V, W)$ does not come from an inner product. In other words, prove that there does not exist an inner product on $\text{Hom}(V, W)$ such that

$$\max\{\|Tv\| : v \in V \text{ and } \|v\| \le 1\} = \sqrt{\langle T, T\rangle}$$

for all $T \in \text{Hom}(V, W)$.

7F.22 Suppose $T \in \text{Hom}(V, W)$. Let $n = \dim V$ and let $s_1 \ge \cdots \ge s_n$ denote the singular values of T. Prove that if $1 \le k \le n$, then

$$\min\{\|T|_U\| : U \text{ is a subspace of } V \text{ with } \dim U = k\} = s_{n-k+1}.$$

7F.23 Suppose $T \in \text{Hom}(V, W)$. Show that T is uniformly continuous with respect to the metrics on V and W that arise from the norms on those spaces (see Exercise 23 in Section 6B).

7F.24 Suppose $T \in \text{End}(V)$ is invertible. Prove that

$$\|T^{-1}\| = \|T\|^{-1} \iff \frac{T}{\|T\|} \text{ is a unitary operator.}$$

7F.25 Fix $u, x \in V$ with $u \neq 0$. Define $T \in \text{End}(V)$ 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 $v \in V$.

7F.26 Suppose $T \in \text{End}(V)$. Prove that T is invertible if and only if there exists a unique unitary operator $S \in \text{End}(V)$ such that $T = S\sqrt{T^*T}$.

7F.27 Suppose $T \in \text{End}(V)$ and $s_1, \ldots, s_n$ are the singular values of T. Let $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ be orthonormal bases 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 all $v \in V$. Define $S \in \text{End}(V)$ by

$$Sv = \langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n.$$

(a) Show that S is unitary and $\|T - S\| = \max\{|s_1 - 1|, \ldots, |s_n - 1|\}$.

(b) Show that if $E \in \text{End}(V)$ is unitary, then $\|T - E\| \geq \|T - S\|$.

This exercise finds a unitary operator S that is as close as possible (among the unitary operators) to a given operator T.

7F.28 Suppose $T \in \text{End}(V)$. Prove that there exists a unitary operator $S \in \text{End}(V)$ such that $T = \sqrt{TT^*}\, S$.

7F.29 Suppose $T \in \text{End}(V)$.

(a) Use the polar decomposition to show that there exists a unitary operator $S \in \text{End}(V)$ such that $TT^* = ST^*TS^*$.

(b) Show how (a) implies that T and T^* have the same singular values.

7F.30 Suppose $T \in \text{End}(V)$, $S \in \text{End}(V)$ is a unitary operator, and $R \in \text{End}(V)$ is a positive operator such that $T = SR$. Prove that $R = \sqrt{T^*T}$.

This exercise shows that if we write T as the product of a unitary operator and a positive operator (as in the polar decomposition 7.93), then the positive operator equals $\sqrt{T^*T}$.

7F.31 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$ is normal. Prove that there exists a unitary operator $S \in \text{End}(V)$ such that $T = S\sqrt{T^*T}$ and such that S and $\sqrt{T^*T}$ both have diagonal matrices with respect to the same orthonormal basis of V.

7F.32 Suppose that $T \in \text{Hom}(V, W)$ and $T \neq 0$. Let $s_1, \ldots, s_m$ denote the positive singular values of T. Show that there exists an orthonormal basis $e_1, \ldots, e_m$ of $(\ker T)^\perp$ such that

$$T\Big(E\Big(\frac{e_1}{s_1}, \ldots, \frac{e_m}{s_m}\Big)\Big)$$

equals the ball in $\text{im}\, T$ of radius 1 centered at 0.

Part 4



Structure

CHAPTER 8	Operators on Complex Vector Spaces	301
CHAPTER 9	Multilinear Algebra and Determinants	337
	Solutions	387
	Colophon	663

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. Some of the results in this chapter apply to both real and complex vector spaces; thus we do not make a standing assumption that $\mathbb{F} = \mathbb{C}$. Also, an inner product does not help with this material, so we return to the general setting of a finite-dimensional vector space.

Even on a finite-dimensional complex vector space, an operator may not have enough eigenvectors to form a basis of the vector space. Thus we will consider the closely related objects called generalized eigenvectors. We will see that for each operator on a finite-dimensional complex vector space, there is a basis of the vector space consisting of generalized eigenvectors of the operator. The generalized eigenspace decomposition then provides a good description of arbitrary operators on a finite-dimensional complex vector space.

Nilpotent operators, which are operators that when raised to some power equal 0, have an important role in these investigations. Nilpotent operators provide a key tool in our proof that every invertible operator on a finite-dimensional complex vector space has a square root and in our approach to Jordan form.

This chapter concludes by defining the trace and proving its key properties.

Standing assumptions for this chapter — Definition

- $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$.
- V denotes a finite-dimensional nonzero vector space over $\mathbb{F}$.

8A 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.

Sequence of increasing null spaces — Result 8.1

Result 8.1 *continued*

Suppose $T \in \text{End}(V)$. Then

$$\{0\} = \ker T^0 \subseteq \ker T^1 \subseteq \cdots \subseteq \ker T^k \subseteq \ker T^{k+1} \subseteq \cdots .$$

Proof. Suppose k is a nonnegative integer and $v \in \ker T^k$. Then $T^k v = 0$, which implies that $T^{k+1}v = T(T^k v) = T(0) = 0$. Thus $v \in \ker T^{k+1}$. Hence $\ker T^k \subseteq \ker T^{k+1}$, as desired. □

For similar results about decreasing sequences of ranges, see Exercises 6, 7, and 8.

The following result states that if two consecutive terms in the sequence of subspaces above are equal, then all later terms in the sequence are equal.

Equality in the sequence of null spaces Result 8.2

Suppose $T \in \text{End}(V)$ and m is a nonnegative integer such that

$$\ker T^m = \ker T^{m+1}.$$

Then

$$\ker T^m = \ker T^{m+1} = \ker T^{m+2} = \ker T^{m+3} = \cdots .$$

Proof. Let k be a positive integer. We want to prove that

$$\ker T^{m+k} = \ker T^{m+k+1}.$$

We already know from 8.1 that $\ker T^{m+k} \subseteq \ker T^{m+k+1}$.

To prove the inclusion in the other direction, suppose $v \in \ker T^{m+k+1}$. Then

$$T^{m+1}(T^k v) = T^{m+k+1}v = 0.$$

Hence

$$T^k v \in \ker T^{m+1} = \ker T^m.$$

Thus $T^{m+k}v = T^m(T^k v) = 0$, which means that $v \in \ker T^{m+k}$. This implies that $\ker T^{m+k+1} \subseteq \ker T^{m+k}$, completing the proof. □

The result above raises the question of whether there exists a nonnegative integer m such that $\ker T^m = \ker T^{m+1}$. The next result shows that this equality holds at least when m equals the dimension of the vector space on which T operates.

Null spaces stop growing Result 8.3

Suppose $T \in \text{End}(V)$. Then

$$\ker T^{\dim V} = \ker T^{\dim V+1} = \ker T^{\dim V+2} = \cdots .$$

Proof. We only need to prove that $\ker T^{\dim V} = \ker T^{\dim V+1}$ (by 8.2). Suppose this is not true. Then, by 8.1 and 8.2, we have

$$\{0\} = \ker T^0 \subsetneq \ker T^1 \subsetneq \cdots \subsetneq \ker T^{\dim V} \subsetneq \ker T^{\dim V+1},$$

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 $\dim \ker T^{\dim V+1} \geq \dim V + 1$, a contradiction because a subspace of V cannot have a larger dimension than $\dim V$. □

It is not true that $V = \ker T \oplus \operatorname{im} T$ for every $T \in \operatorname{End}(V)$. However, the next result can be a useful substitute.

Result 8.4

V is the direct sum of $\ker T^{\dim V}$ and $\operatorname{im} T^{\dim V}$
Suppose $T \in \operatorname{End}(V)$. Then

$$V = \ker T^{\dim V} \oplus \operatorname{im} T^{\dim V}.$$

Proof. Let $n = \dim V$. First we show that

$$(\ker T^n) \cap (\operatorname{im} T^n) = \{0\}. \tag{8.5}$$

Suppose $v \in (\ker T^n) \cap (\operatorname{im} T^n)$. 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 8.3). Thus $v = T^n u = 0$, completing the proof of 8.5.

Now 8.5 implies that $\ker T^n + \operatorname{im} T^n$ is a direct sum (by 1.46). Also,

$$\dim(\ker T^n \oplus \operatorname{im} T^n) = \dim \ker T^n + \dim \operatorname{im} T^n = \dim V,$$

where the first equality above comes from 3.94 and the second equality comes from the fundamental theorem of linear maps (3.21). The equation above implies that $\ker T^n \oplus \operatorname{im} T^n = V$ (see 2.39), as desired. □

For an improvement of the result above, see Exercise 19.

Example 8.6 $\mathbb{F}^3 = \ker T^3 \oplus \operatorname{im} T^3$ **for** $T \in \operatorname{End}(\mathbb{F}^3)$

Suppose $T \in \operatorname{End}(\mathbb{F}^3)$ is defined by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

Then $\ker T = \{(z_1, 0, 0) : z_1 \in \mathbb{F}\}$ and $\operatorname{im} T = \{(z_1, 0, z_3) : z_1, z_3 \in \mathbb{F}\}$. Thus $\ker T \cap \operatorname{im} T \neq \{0\}$. Hence $\ker T + \operatorname{im} T$ is not a direct sum. Also note that $\ker T + \operatorname{im} T \neq \mathbb{F}^3$. However, we have $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$. Thus we see that

$$\ker T^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbb{F}\} \quad \text{and} \quad \operatorname{im} T^3 = \{(0, 0, z_3) : z_3 \in \mathbb{F}\}.$$

Hence $\mathbb{F}^3 = \ker T^3 \oplus \operatorname{im} T^3$, as expected by 8.4.

Generalized Eigenvectors

Some operators do not have enough eigenvectors to lead to good descriptions of their behavior. 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 $T \in \text{End}(V)$. We seek to describe T by finding a "nice" direct sum decomposition

$$V = V_1 \oplus \cdots \oplus V_n,$$

where each V_k is a subspace of V invariant under T. The simplest possible nonzero invariant subspaces are one-dimensional. A decomposition as above in which each V_k is a one-dimensional subspace of V invariant under T is possible if and only if V has a basis consisting of eigenvectors of T (see 5.55). This happens if and only if V has an eigenspace decomposition

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T), \tag{8.7}$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T (see 5.55).

The spectral theorem in the previous chapter shows that if V is an inner product space, then a decomposition of the form 8.7 holds for every self-adjoint operator if $\mathbb{F} = \mathbb{R}$ and for every normal operator if $\mathbb{F} = \mathbb{C}$ because operators of those types have enough eigenvectors to form a basis of V (see 7.29 and 7.31).

However, a decomposition of the form 8.7 may not hold for more general operators, even on a complex vector space. An example was given by the operator in 5.57, which does not have enough eigenvectors for 8.7 to hold. Generalized eigenvectors and generalized eigenspaces, which we now introduce, will remedy this situation.

Generalized eigenvector — **Definition 8.8**

Suppose $T \in \text{End}(V)$ 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

$$(T - \lambda I)^k v = 0$$

for some positive integer k.

A nonzero vector $v \in V$ is a generalized eigenvector of T corresponding to λ if and only if

$$(T - \lambda I)^{\dim V} v = 0,$$

as follows from applying 8.1 and 8.3 to the operator $T - \lambda I$.

Generalized eigenvalues are not defined because doing so would not lead to anything new. Reason: if $(T - \lambda I)^k$ is not injective for some positive integer k, then $T - \lambda I$ is not injective, and hence λ is an eigenvalue of T.

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.

A basis of generalized eigenvectors — Result 8.9

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.

Proof. Let $n = \dim V$. We will use induction on n. To get started, note that the desired result holds if $n = 1$ because then every nonzero vector in V is an eigenvector of T.

Now suppose $n > 1$ and the desired result holds for all smaller values of $\dim V$. Let λ be an eigenvalue of T. Applying 8.4 to $T - \lambda I$ shows that

$$V = \ker(T - \lambda I)^n \oplus \text{im}(T - \lambda I)^n.$$

This step is where we use the hypothesis that $\mathbb{F} = \mathbb{C}$, because if $\mathbb{F} = \mathbb{R}$ then T may not have any eigenvalues.

If $\ker(T - \lambda I)^n = V$, then every nonzero vector in V is a generalized eigenvector of T, and thus in this case there is a basis of V consisting of generalized eigenvectors of T. Hence we can assume that $\ker(T - \lambda I)^n \neq V$, which implies that $\text{im}(T - \lambda I)^n \neq \{0\}$.

Also, $\ker(T - \lambda I)^n \neq \{0\}$, because λ is an eigenvalue of T. Thus we have

$$0 < \dim \text{im}(T - \lambda I)^n < n.$$

Furthermore, $\text{im}(T - \lambda I)^n$ is invariant under T [by 5.18 with $p(z) = (z - \lambda)^n$]. Let $S \in \text{End}\big(\text{im}(T - \lambda I)^n\big)$ equal T restricted to $\text{im}(T - \lambda I)^n$. Our induction hypothesis applied to the operator S implies that there is a basis of $\text{im}(T - \lambda I)^n$ consisting of generalized eigenvectors of S, which of course are generalized eigenvectors of T. Adjoining that basis of $\text{im}(T - \lambda I)^n$ to a basis of $\ker(T - \lambda I)^n$ gives a basis of V consisting of generalized eigenvectors of T. □

If $\mathbb{F} = \mathbb{R}$ and $\dim V > 1$, then some operators on V have the property that there exists a basis of V consisting of generalized eigenvectors of the operator, and (unlike what happens when $\mathbb{F} = \mathbb{C}$) other operators do not have this property. See Exercise 11 for a necessary and sufficient condition that determines whether an operator has this property.

Example 8.10 **Generalized eigenvectors of an operator on $\mathbb{C}^3$**

Define $T \in \text{End}(\mathbb{C}^3)$ by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3)$$

for each $(z_1, z_2, z_3) \in \mathbb{C}^3$. A routine use of the definition of eigenvalue shows that the eigenvalues of T are 0 and 5. Furthermore, the eigenvectors corresponding to the eigenvalue 0 are the nonzero vectors of the form $(z_1, 0, 0)$, and the eigenvectors corresponding to the eigenvalue 5 are the nonzero vectors of the form $(0, 0, z_3)$. Hence this operator does not have enough eigenvectors to span its domain $\mathbb{C}^3$.

We compute that $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$. Thus 8.1 and 8.3 imply that the generalized eigenvectors of T corresponding to the eigenvalue 0 are the nonzero vectors of the form $(z_1, z_2, 0)$.

We also have $(T - 5I)^3(z_1, z_2, z_3) = (-125z_1 + 300z_2, -125z_2, 0)$. Thus the

Example 8.10
continued

generalized eigenvectors of T corresponding to the eigenvalue 5 are the nonzero vectors of the form $(0, 0, z_3)$.

The paragraphs above show that each of the standard basis vectors of $\mathbb{C}^3$ is a generalized eigenvector of T. Thus $\mathbb{C}^3$ indeed has a basis consisting of generalized eigenvectors of T, as promised by 8.9.

If v is an eigenvector of $T \in \text{End}(V)$, then the corresponding eigenvalue λ is uniquely determined by the equation $Tv = \lambda v$, which can be satisfied by only one $\lambda \in \mathbb{F}$ (because $v \neq 0$). However, if v is a generalized eigenvector of T, then it is not obvious that the equation $(T - \lambda I)^{\dim V} v = 0$ can be satisfied by only one $\lambda \in \mathbb{F}$. Fortunately, the next result tells us that all is well on this issue.

Generalized eigenvector corresponds to a unique eigenvalue Result 8.11

Suppose $T \in \text{End}(V)$. Then each generalized eigenvector of T corresponds to only one eigenvalue of T.

Proof. Suppose $v \in V$ is a generalized eigenvector of T corresponding to eigenvalues α and λ of T. Let m be the smallest positive integer such that $(T - \alpha I)^m v = 0$. Let $n = \dim V$. Then

$$\begin{aligned} 0 &= (T - \lambda I)^n v \\ &= \big((T - \alpha I) + (\alpha - \lambda) I\big)^n v \\ &= \sum_{k=0}^{n} b_k (\alpha - \lambda)^{n-k} (T - \alpha I)^k v, \end{aligned}$$

where $b_0 = 1$ and the values of the other binomial coefficients b_k do not matter. Apply the operator $(T - \alpha I)^{m-1}$ to both sides of the equation above, getting

$$0 = (\alpha - \lambda)^n (T - \alpha I)^{m-1} v.$$

Because $(T - \alpha I)^{m-1} v \neq 0$, the equation above implies that $\alpha = \lambda$, as desired. □

We saw earlier (5.11) that eigenvectors corresponding to distinct eigenvalues are linearly independent. Now we prove a similar result for generalized eigenvectors, with a proof that roughly follows the pattern of the proof of that earlier result.

Linearly independent generalized eigenvectors Result 8.12

Suppose that $T \in \text{End}(V)$. Then every list of generalized eigenvectors of T corresponding to distinct eigenvalues of T is linearly independent.

Proof. Suppose the desired result is false. Then there exists a smallest positive integer m such that there exists a linearly dependent list $v_1, \ldots, v_m$ of generalized eigenvectors of T corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ of T (note that $m \geq 2$ because

a generalized eigenvector is, by definition, nonzero). Thus there exist $a_1, \dots, a_m \in \mathbb{F}$, none of which are 0 (because of the minimality of m), such that

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

Let $n = \dim V$. Apply $(T - \lambda_m I)^n$ to both sides of the equation above, getting

$$a_1(T - \lambda_m I)^n v_1 + \cdots + a_{m-1}(T - \lambda_m I)^n v_{m-1} = 0. \tag{8.13}$$

Suppose $k \in \{1, \dots, m-1\}$. Then

$$(T - \lambda_m I)^n v_k \neq 0$$

because otherwise v_k would be a generalized eigenvector of T corresponding to the distinct eigenvalues λ_k and λ_m, which would contradict 8.11. However,

$$(T - \lambda_k I)^n \big((T - \lambda_m I)^n v_k\big) = (T - \lambda_m I)^n \big((T - \lambda_k I)^n v_k\big) = 0.$$

Thus the last two displayed equations show that $(T - \lambda_m I)^n v_k$ is a generalized eigenvector of T corresponding to the eigenvalue λ_k. Hence

$$(T - \lambda_m I)^n v_1, \dots, (T - \lambda_m I)^n v_{m-1}$$

is a linearly dependent list (by 8.13) of $m-1$ generalized eigenvectors corresponding to distinct eigenvalues, contradicting the minimality of m. This contradiction completes the proof. □

Nilpotent Operators

Nilpotent — Definition 8.14

An operator is called *nilpotent* if some power of it equals 0.

Thus an operator $T \in \text{End}(V)$ is nilpotent if and only if every nonzero vector in V is a generalized eigenvector of T corresponding to the eigenvalue 0.

Example 8.15 **Nilpotent operators**

(a) The operator $T \in \text{End}(\mathbb{F}^4)$ defined by

$$T(z_1, z_2, z_3, z_4) = (0, 0, z_1, z_2)$$

is nilpotent because $T^2 = 0$.

Example 8.15
continued

(b) The operator on $\mathbb{F}^3$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} -3 & 9 & 0 \\ -7 & 9 & 6 \\ 4 & 0 & -6 \end{pmatrix}$$

is nilpotent, as can be shown by cubing the matrix above to get the zero matrix.

(c) The operator of differentiation on $\mathbb{R}[X]_{\le m}$ is nilpotent because the $(m+1)^{\text{th}}$ 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 **potens** means having power. Thus **nilpotent** literally means having a power that is zero.

The next result shows that when raising a nilpotent operator to a power, we never need to use a power higher than the dimension of the space. For a slightly stronger result, see Exercise 18.

Nilpotent operator raised to dimension of domain is 0 — Result 8.16

Suppose $T \in \text{End}(V)$ is nilpotent. Then $T^{\dim V} = 0$.

Proof. Because T is nilpotent, there exists a positive integer k such that $T^k = 0$. Thus $\ker T^k = V$. Now 8.1 and 8.3 imply that $\ker T^{\dim V} = V$. Thus $T^{\dim V} = 0$. □

Eigenvalues of nilpotent operator — Result 8.17

Suppose $T \in \text{End}(V)$.

(a) If T is nilpotent, then 0 is an eigenvalue of T and T has no other eigenvalues.

(b) If $\mathbb{F} = \mathbb{C}$ and 0 is the only eigenvalue of T, then T is nilpotent.

Proof.

(a) To prove (a), suppose T is nilpotent. Hence there is a positive integer m such that $T^m = 0$. This implies that T is not injective. Thus 0 is an eigenvalue of T.

To show that T has no other eigenvalues, suppose λ is an eigenvalue of T. Then there exists a nonzero vector $v \in V$ such that

$$\lambda v = Tv.$$

Repeatedly applying T to both sides of this equation shows that

$$\lambda^m v = T^m v = 0.$$

Thus $\lambda = 0$, as desired.

(b) Suppose $\mathbb{F} = \mathbb{C}$ and 0 is the only eigenvalue of T. By 5.27(b), the minimal polynomial of T equals z^m for some positive integer m. Thus $T^m = 0$. Hence T is nilpotent.

□

Exercise 23 shows that the hypothesis that $\mathbb{F} = \mathbb{C}$ cannot be deleted in (b) of the result above.

Given an operator on V, we want to find a basis of V such that the matrix of the operator with respect to this basis is as simple as possible, meaning that the matrix contains many 0's. The next result shows that if T is nilpotent, then we can choose a basis of V such that the matrix of T with respect to this basis has more than half of its entries equal to 0. Later in this chapter we will do even better.

Minimal polynomial and upper-triangular matrix of nilpotent operator Result 8.18

Suppose $T \in \text{End}(V)$. Then the following are equivalent.

(a) T is nilpotent.

(b) The minimal polynomial of T is z^m for some positive integer m.

(c) There is a basis of V with respect to which the matrix of T has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix},$$

where all entries on and below the diagonal equal 0.

Proof. Suppose (a) holds, so T is nilpotent. Thus there exists a positive integer n such that $T^n = 0$. Now 5.29 implies that z^n is a polynomial multiple of the minimal polynomial of T. Thus the minimal polynomial of T is z^m for some positive integer m, proving that (a) implies (b).

Now suppose (b) holds, so the minimal polynomial of T is z^m for some positive integer m. This implies, by 5.27(a), that 0 (which is the only zero of z^m) is the only eigenvalue of T. This further implies, by 5.44, that there is a basis of V with respect to which the matrix of T is upper triangular. This also implies, by 5.41, that all entries on the diagonal of this matrix are 0, proving that (b) implies (c).

Now suppose (c) holds. Then 5.40 implies that $T^{\dim V} = 0$. Thus T is nilpotent, proving that (c) implies (a). □

Exercises

8A.1 Suppose $T \in \text{End}(V)$. Prove that if $\dim \ker T^4 = 8$ and $\dim \ker T^6 = 9$, then $\dim \ker T^m = 9$ for all integers $m \geq 5$.

8A.2 Suppose $T \in \text{End}(V)$, m is a positive integer, $v \in V$, and $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that $v, Tv, T^2v, \ldots, T^{m-1}v$ is linearly independent.

The result in this exercise is used in the proof of 8.45.

8A.3 Suppose $T \in \text{End}(V)$. Prove that

$$V = \ker T \oplus \text{im}\, T \iff \ker T^2 = \ker T.$$

8A.4 Suppose $T \in \text{End}(V)$, $\lambda \in \mathbb{F}$, and m is a positive integer such that the minimal polynomial of T is a polynomial multiple of $(z - \lambda)^m$. Prove that

$$\dim \ker (T - \lambda I)^m \geq m.$$

8A.5 Suppose $T \in \text{End}(V)$ and m is a positive integer. Prove that

$$\dim \ker T^m \leq m \dim \ker T.$$

Hint: Exercise 21 in Section 3B may be useful.

8A.6 Suppose $T \in \text{End}(V)$. Show that

$$V = \text{im}\, T^0 \supseteq \text{im}\, T^1 \supseteq \cdots \supseteq \text{im}\, T^k \supseteq \text{im}\, T^{k+1} \supseteq \cdots.$$

8A.7 Suppose $T \in \text{End}(V)$ and m is a nonnegative integer such that

$$\text{im}\, T^m = \text{im}\, T^{m+1}.$$

Prove that $\text{im}\, T^k = \text{im}\, T^m$ for all $k > m$.

8A.8 Suppose $T \in \text{End}(V)$. Prove that

$$\text{im}\, T^{\dim V} = \text{im}\, T^{\dim V + 1} = \text{im}\, T^{\dim V + 2} = \cdots.$$

8A.9 Suppose $T \in \text{End}(V)$ and m is a nonnegative integer. Prove that

$$\ker T^m = \ker T^{m+1} \iff \text{im}\, T^m = \text{im}\, T^{m+1}.$$

8A.10 Define $T \in \text{End}(\mathbb{C}^2)$ by $T(w, z) = (z, 0)$. Find all generalized eigenvectors of T.

8A.11 Suppose that $T \in \text{End}(V)$. Prove that there is a basis of V consisting of generalized eigenvectors of T if and only if the minimal polynomial of T equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.

Assume $\mathbb{F} = \mathbb{R}$ because the case $\mathbb{F} = \mathbb{C}$ follows from 5.27(b) and 8.9.

This exercise states that the condition for there to be a basis of V consisting of generalized eigenvectors of T is the same as the condition for there to be a basis with respect to which T has an upper-triangular matrix (see 5.44).

***Caution:** If T has an upper-triangular matrix with respect to a basis $v_1, \ldots, v_n$ of V, then v_1 is an eigenvector of T but it is not necessarily true that $v_2, \ldots, v_n$ are generalized eigenvectors of T.*

8A.12 Suppose $T \in \text{End}(V)$ is such that every nonzero vector in V is a generalized eigenvector of T. Prove that there exists $\lambda \in \mathbb{F}$ such that $T - \lambda I$ is nilpotent.

8A.13 Suppose $S, T \in \text{End}(V)$ and ST is nilpotent. Prove that TS is nilpotent.

8A.14 Suppose $T \in \text{End}(V)$ is nilpotent and $T \neq 0$. Prove T is not diagonalizable.

8A.15 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that T is diagonalizable if and only if every generalized eigenvector of T is an eigenvector of T.

For $\mathbb{F} = \mathbb{C}$, this exercise adds another equivalence to the list of conditions for diagonalizability in 5.55.

8A.16

(a) Give an example of nilpotent operators S, T on the same vector space such that neither $S + T$ nor ST is nilpotent.

(b) Suppose $S, T \in \text{End}(V)$ are nilpotent and $ST = TS$. Prove that $S + T$ and ST are nilpotent.

8A.17 Suppose $T \in \text{End}(V)$ is nilpotent and m is a positive integer such that $T^m = 0$.

(a) Prove that $I - T$ is invertible and that $(I - T)^{-1} = I + T + \cdots + T^{m-1}$.

(b) Explain how you would guess the formula above.

8A.18 Suppose $T \in \text{End}(V)$ is nilpotent. Prove that $T^{1+\dim \text{im}\, T} = 0$.

If $\dim \text{im}\, T < \dim V - 1$, then this exercise improves 8.16.

8A.19 Suppose $T \in \text{End}(V)$ is not nilpotent. Show that

$$V = \ker T^{\dim V - 1} \oplus \text{im}\, T^{\dim V - 1}.$$

For operators that are not nilpotent, this exercise improves 8.4.

8A.20 Suppose V is an inner product space and $T \in \text{End}(V)$ is normal and nilpotent. Prove that $T = 0$.

8A.21 Suppose $T \in \text{End}(V)$ is such that $\ker T^{\dim V - 1} \neq \ker T^{\dim V}$. Prove that T is nilpotent and that $\dim \ker T^k = k$ for every integer k with $0 \leq k \leq \dim V$.

8A.22 Suppose $T \in \text{End}(\mathbb{C}^5)$ is such that $\text{im}\, T^4 \neq \text{im}\, T^5$. Prove that T is nilpotent.

8A.23 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.

This exercise shows that the implication (b) $\Longrightarrow$ (a) *in 8.17 does not hold without the hypothesis that* $\mathbb{F} = \mathbb{C}$.

8A.24 For each item in Example 8.15, find a basis of the domain vector space such that the matrix of the nilpotent operator with respect to that basis has the upper-triangular form promised by 8.18(c).

8A.25 Suppose that V is an inner product space and $T \in \text{End}(V)$ is nilpotent. Show that there is an orthonormal basis of V with respect to which the matrix of T has the upper-triangular form promised by 8.18(c).

8B Generalized Eigenspace Decomposition

Generalized Eigenspaces

Generalized eigenspace, $G(\lambda, T)$ — **Definition 8.19**

Suppose $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$. The *generalized eigenspace* of T corresponding to λ, denoted by $G(\lambda, T)$, is defined by

$$G(\lambda, T) = \{v \in V : (T - \lambda I)^k v = 0 \text{ for some positive integer } k\}.$$

Thus $G(\lambda, T)$ is the set of generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Because every eigenvector of T is a generalized eigenvector of T (take $k = 1$ in the definition of generalized eigenvector), each eigenspace is contained in the corresponding generalized eigenspace. In other words, if $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$, then $E(\lambda, T) \subseteq G(\lambda, T)$.

The next result implies that if $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$, then the generalized eigenspace $G(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V).

Description of generalized eigenspaces **Result 8.20**

Suppose $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$. Then $G(\lambda, T) = \ker(T - \lambda I)^{\dim V}$.

Proof. Suppose $v \in \ker(T - \lambda I)^{\dim V}$. The definitions imply $v \in G(\lambda, T)$. Thus $G(\lambda, T) \supseteq \ker(T - \lambda I)^{\dim V}$.

Conversely, suppose $v \in G(\lambda, T)$. Thus there is a positive integer k such that $v \in \ker(T-\lambda I)^k$. From 8.1 and 8.3 (with $T-\lambda I$ replacing T), we get $v \in \ker(T-\lambda I)^{\dim V}$. Thus $G(\lambda, T) \subseteq \ker(T - \lambda I)^{\dim V}$, completing the proof. □

Example 8.21 **Generalized eigenspaces of an operator on $\mathbb{C}^3$**

Define $T \in \text{End}(\mathbb{C}^3)$ by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3).$$

In Example 8.10, we saw that the eigenvalues of T are 0 and 5, and we found the corresponding sets of generalized eigenvectors. Taking the union of those sets with $\{0\}$, we have

$$G(0, T) = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbb{C}\} \quad \text{and} \quad G(5, T) = \{(0, 0, z_3) : z_3 \in \mathbb{C}\}.$$

Note that $\mathbb{C}^3 = G(0, T) \oplus G(5, T)$.

In Example 8.21, the domain space $\mathbb{C}^3$ is the direct sum of the generalized eigenspaces of the operator T in that example. Our next result shows that this behavior holds in general. Specifically, the following major result shows that if $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$, then V is the direct sum of the generalized eigenspaces of T, each of which is invariant under T and on which T is a nilpotent operator plus a scalar multiple of the identity. Thus the next result achieves our goal of decomposing V into invariant subspaces on which T has a known behavior.

As we will see, the proof follows from putting together what we have learned about generalized eigenspaces and then using our result that for each operator $T \in \text{End}(V)$, there exists a basis of V consisting of generalized eigenvectors of T.

Generalized eigenspace decomposition **Result 8.22**

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. Then

(a) $G(\lambda_k, T)$ is invariant under T for each $k = 1, \ldots, m$;

(b) $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent for each $k = 1, \ldots, m$;

(c) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$.

Proof.

(a) Suppose $k \in \{1, \dots, m\}$. Then 8.20 shows that

$$G(\lambda_k, T) = \ker(T - \lambda_k I)^{\dim V}.$$

Thus 5.18, with $p(z) = (z - \lambda_k)^{\dim V}$, implies that $G(\lambda_k, T)$ is invariant under T, proving (a).

(b) Suppose $k \in \{1, \dots, m\}$. If $v \in G(\lambda_k, T)$, then $(T - \lambda_k I)^{\dim V} v = 0$ (by 8.20). Thus $\big((T - \lambda_k I)|_{G(\lambda_k, T)}\big)^{\dim V} = 0$. Hence $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent, proving (b).

(c) To show that $G(\lambda_1, T) + \cdots + G(\lambda_m, T)$ is a direct sum, suppose

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

where each v_k is in $G(\lambda_k, T)$. Because generalized eigenvectors of T corresponding to distinct eigenvalues are linearly independent (by 8.12), this implies that each v_k equals 0. Thus $G(\lambda_1, T) + \cdots + G(\lambda_m, T)$ is a direct sum (by 1.45).

Finally, each vector in V can be written as a finite sum of generalized eigenvectors of T (by 8.9). Thus

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

proving (c). □

For the analogous result when $\mathbb{F} = \mathbb{R}$, see Exercise 8.

Multiplicity of an Eigenvalue

If V is a complex vector space and $T \in \text{End}(V)$, then the decomposition of V provided by the generalized eigenspace decomposition (8.22) can be a powerful tool. The dimensions of the subspaces involved in this decomposition are sufficiently important to get a name, which is given in the next definition.

Definition 8.23

Multiplicity

- Suppose $T \in \text{End}(V)$. 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

$$\dim \ker(T - \lambda I)^{\dim V}.$$

The second bullet point above holds because $G(\lambda, T) = \ker(T - \lambda I)^{\dim V}$ (see 8.20).

Example 8.24 **Multiplicity of each eigenvalue of an operator**

Example 8.24 *continued*

Suppose $T \in \text{End}(\mathbb{C}^3)$ 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 the diagonal entries 6 and 7, as follows from 5.41. You can verify that the generalized eigenspaces of T are as follows:

$$G(6, T) = \text{span}\big((1, 0, 0), (0, 1, 0)\big) \quad \text{and} \quad G(7, T) = \text{span}\big((10, 2, 1)\big).$$

Thus the eigenvalue 6 has multiplicity 2 and the eigenvalue 7 has multiplicity 1. The direct sum $\mathbb{C}^3 = G(6, T) \oplus G(7, T)$ is the generalized eigenspace decomposition promised by 8.22. A basis of $\mathbb{C}^3$ consisting of generalized eigenvectors of T, as promised by 8.9, is $(1, 0, 0), (0, 1, 0), (10, 2, 1)$. There does not exist a basis of $\mathbb{C}^3$ consisting of eigenvectors of this operator.

In this example, the multiplicity of each eigenvalue equals the number of times that eigenvalue appears on the diagonal of an upper-triangular matrix representing the operator. This behavior always happens, as we will see in 8.31.

In the example above, 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 holds for all operators on finite-dimensional complex vector spaces.

Result 8.25

Sum of the multiplicities equals $\dim V$

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Then the sum of the multiplicities of all eigenvalues of T equals $\dim V$.

Proof. The desired result follows from the generalized eigenspace decomposition (8.22) and the formula for the dimension of a direct sum (see 3.94). □

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 $T \in \text{End}(V)$ and λ is an eigenvalue of T, then

$$\begin{aligned} &\text{algebraic multiplicity of } \lambda = \dim \ker(T - \lambda I)^{\dim V} = \dim G(\lambda, T), \\ &\text{geometric multiplicity of } \lambda = \dim \ker(T - \lambda I) = \dim E(\lambda, T). \end{aligned}$$

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; 9.62 implies that these definitions are equivalent.

If V is an inner product space, $T \in \text{End}(V)$ is normal, and λ is an eigenvalue of T, then the algebraic multiplicity of λ equals the geometric multiplicity of λ, as

can be seen from applying Exercise 27 in Section 7A to the normal operator $T - \lambda I$. As a special case, the singular values of $S \in \text{Hom}(V, W)$ (here V and W are both finite-dimensional inner product spaces) depend on the multiplicities (either algebraic or geometric) of the eigenvalues of the self-adjoint operator S^*S.

The next definition associates a monic polynomial with each operator on a finite-dimensional complex vector space.

Characteristic polynomial — Definition 8.26

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \ldots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the *characteristic polynomial* of T.

Example 8.27 **The characteristic polynomial of an operator**

Suppose $T \in \text{End}(\mathbb{C}^3)$ is defined as in Example 8.24. Because the eigenvalues of T are 6, with multiplicity 2, and 7, with multiplicity 1, we see that the characteristic polynomial of T is $(z - 6)^2(z - 7)$.

Degree and zeros of characteristic polynomial — Result 8.28

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. 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. Our result about the sum of the multiplicities (8.25) implies (a). The definition of the characteristic polynomial implies (b). □

Most texts define the characteristic polynomial using determinants (the two definitions are equivalent by 9.62). The approach taken here, which is considerably simpler, leads to the following nice proof of the Cayley–Hamilton theorem.

Cayley–Hamilton theorem — Result 8.29

Suppose $\mathbb{F} = \mathbb{C}$, $T \in \text{End}(V)$, and q is the characteristic polynomial of T. Then $q(T) = 0$.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, and let $d_k = \dim G(\lambda_k, T)$. For each $k \in \{1, \ldots, m\}$, we know that $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent. Thus we have

$$(T - \lambda_k I)^{d_k}|_{G(\lambda_k, T)} = 0$$

Arthur Cayley (1821–1895) published three mathematics papers before completing his undergraduate degree.

(by 8.16) for each $k \in \{1, \dots, m\}$.

The generalized eigenspace decomposition (8.22) states that every vector in V is a sum of vectors in $G(\lambda_1, T), \dots, G(\lambda_m, T)$. Thus to prove that $q(T) = 0$, we only need to show that $q(T)|_{G(\lambda_k, T)} = 0$ for each k.

Fix $k \in \{1, \dots, m\}$. 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 $(T - \lambda_k I)^{d_k}$ to be the last term in the expression on the right. Because $(T - \lambda_k I)^{d_k}|_{G(\lambda_k, T)} = 0$, we have $q(T)|_{G(\lambda_k, T)} = 0$, as desired. □

The next result implies that if the minimal polynomial of an operator $T \in \text{End}(V)$ has degree $\dim V$ (as happens almost always—see the paragraphs following 5.24), then the characteristic polynomial of T equals the minimal polynomial of T.

Characteristic polynomial is a multiple of minimal polynomial **Result 8.30**

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. 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 (8.29) and 5.29. □

Now we can prove that the result suggested by Example 8.24 holds for all operators on finite-dimensional complex vector spaces.

Multiplicity of an eigenvalue equals number of times on diagonal **Result 8.31**

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Suppose $v_1, \dots, v_n$ is a basis of V such that $\text{Mat}\big(T, (v_1, \dots, v_n)\big)$ is upper triangular. Then the number of times that each eigenvalue λ of T appears on the diagonal of $\text{Mat}\big(T, (v_1, \dots, v_n)\big)$ equals the multiplicity of λ as an eigenvalue of T.

Proof. Let $A = \text{Mat}\big(T, (v_1, \dots, v_n)\big)$. Thus A is an upper-triangular matrix. Let $\lambda_1, \dots, \lambda_n$ denote the entries on the diagonal of A. Thus for each $k \in \{1, \dots, n\}$, we have

$$Tv_k = u_k + \lambda_k v_k$$

for some $u_k \in \text{span}(v_1, \dots, v_{k-1})$. Hence if $k \in \{1, \dots, n\}$ and $\lambda_k \neq 0$, then Tv_k is not a linear combination of $Tv_1, \dots, Tv_{k-1}$. The linear dependence lemma (2.19) now implies that the list of those Tv_k such that $\lambda_k \neq 0$ is linearly independent.

Let d denote the number of indices $k \in \{1, \dots, n\}$ such that $\lambda_k = 0$. The conclusion of the previous paragraph implies that

$$\dim \text{im}\, T \geq n - d.$$

Because $n = \dim V = \dim \ker T + \dim \text{im}\, T$, the inequality above implies that

$$\dim \ker T \leq d. \qquad 8.32$$

The matrix of the operator T^n with respect to the basis $v_1, \dots, v_n$ is the upper-triangular matrix A^n, which has diagonal entries $\lambda_1^n, \dots, \lambda_n^n$ [see Exercise 2(b) in Section 5C]. Because $\lambda_k^n = 0$ if and only if $\lambda_k = 0$, the number of times that 0 appears on the diagonal of A^n equals d. Thus applying 8.32 with T replaced with T^n, we have

$$\dim \ker T^n \le d. \tag{8.33}$$

For λ an eigenvalue of T, let m_λ denote the multiplicity of λ as an eigenvalue of T and let d_λ denote the number of times that λ appears on the diagonal of A. Replacing T in 8.33 with $T - \lambda I$, we see that

$$m_\lambda \le d_\lambda \tag{8.34}$$

for each eigenvalue λ of T. The sum of the multiplicities m_λ over all eigenvalues λ of T equals n, the dimension of V (by 8.25). The sum of the numbers d_λ over all eigenvalues λ of T also equals n, because the diagonal of A has length n.

Thus summing both sides of 8.34 over all eigenvalues λ of T produces an equality. Hence 8.34 must actually be an equality for each eigenvalue λ of T. Thus the multiplicity of λ as an eigenvalue of T equals the number of times that λ appears on the diagonal of A, as desired. □

Block Diagonal Matrices

To interpret our results in matrix form, we make the following definition, generalizing the notion of a diagonal matrix. If each matrix A_k in the definition below is a 1-by-1 matrix, then we actually have a diagonal matrix.

Often we can understand a matrix better by thinking of it as composed of smaller matrices.

Block diagonal matrix — Definition 8.35

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 other entries of the matrix equal 0.

Example 8.36 **A block diagonal matrix**

Example 8.36 *continued*

The 5-by-5 matrix

$$A = \begin{pmatrix} (4) & 0 & 0 & 0 & 0 \\ 0 & \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix} & & 0 & 0 \\ 0 & & & 0 & 0 \\ 0 & 0 & 0 & \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix} & \\ 0 & 0 & 0 & & \end{pmatrix}$$

is a block diagonal matrix with

$$A = \begin{pmatrix} A_1 & & 0 \\ & A_2 & \\ 0 & & A_3 \end{pmatrix},$$

where

$$A_1 = (4), \quad A_2 = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 7 \\ 0 & 1 \end{pmatrix}.$$

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 example above, each of A_1, A_2, A_3 is an upper-triangular matrix whose diagonal entries are all equal. The next result shows that with respect to an appropriate basis, every operator on a finite-dimensional complex vector space has a matrix of this form. Note that this result gives us many more zeros in the matrix than are needed to make it upper triangular.

Result 8.37

Block diagonal matrix with upper-triangular blocks

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. 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_k is a d_k-by-d_k upper-triangular matrix of the form

$$A_k = \begin{pmatrix} \lambda_k & & * \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}.$$

Proof. Each $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent (see 8.22). For each k, choose a basis of $G(\lambda_k, T)$, which is a vector space of dimension d_k, such that the matrix of $(T - \lambda_k I)|_{G(\lambda_k, T)}$ with respect to this basis is as in 8.18(c). Thus with respect to this basis, the matrix of $T|_{G(\lambda_k, T)}$, which equals $(T - \lambda_k I)|_{G(\lambda_k, T)} + \lambda_k I|_{G(\lambda_k, T)}$, looks like the desired form shown above for A_k.

The generalized eigenspace decomposition (8.22) shows that putting together the bases of the $G(\lambda_k, T)$'s chosen above gives a basis of V. The matrix of T with respect to this basis has the desired form. □

Example 8.38 **Block diagonal matrix via generalized eigenvectors**

Let $T \in \text{End}(\mathbb{C}^3)$ be 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 8.37.

As we saw in Example 8.24, the eigenvalues of T are 6 and 7; also,

$$G(6, T) = \text{span}\big((1, 0, 0), (0, 1, 0)\big) \quad \text{and} \quad G(7, T) = \text{span}\big((10, 2, 1)\big).$$

We also saw that a basis of $\mathbb{C}^3$ 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

$$\begin{pmatrix} \begin{pmatrix} 6 & 3 \\ 0 & 6 \end{pmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & (7) \end{pmatrix},$$

which is a matrix of the block diagonal form promised by 8.37.

Exercises

8B.1 Define $T \in \text{End}(\mathbb{C}^2)$ by $T(w, z) = (-z, w)$. Find the generalized eigenspaces corresponding to the distinct eigenvalues of T.

8B.2 Suppose $T \in \text{End}(V)$ is invertible. Prove that $G(\lambda, T) = G(1/\lambda, T^{-1})$ for every $\lambda \in \mathbb{F}$ with $\lambda \neq 0$.

8B.3 Suppose $T \in \text{End}(V)$. Suppose $S \in \text{End}(V)$ is invertible. Prove that T and $S^{-1}TS$ have the same eigenvalues with the same multiplicities.

8B.4 Suppose $\dim V \geq 2$ and $T \in \text{End}(V)$ is such that $\ker T^{\dim V - 2} \neq \ker T^{\dim V - 1}$. Prove that T has at most two distinct eigenvalues.

8B.5 Suppose $T \in \text{End}(V)$ and 3 and 8 are eigenvalues of T. Let $n = \dim V$. Prove that $V = (\ker T^{n-2}) \oplus (\text{im } T^{n-2})$.

8B.6 Suppose $T \in \text{End}(V)$ and λ is an eigenvalue of T. Explain why the exponent of $z - \lambda$ in the factorization of the minimal polynomial of T is the smallest positive integer m such that $(T - \lambda I)^m|_{G(\lambda,T)} = 0$.

8B.7 Suppose $T \in \text{End}(V)$ and λ is an eigenvalue of T with multiplicity d. Prove that $G(\lambda, T) = \ker(T - \lambda I)^d$.

If $d < \dim V$, then this exercise improves 8.20.

8B.8 Suppose $T \in \text{End}(V)$ and $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T. Prove that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$$

if and only if the minimal polynomial of T equals $(z - \lambda_1)^{k_1} \cdots (z - \lambda_m)^{k_m}$ for some positive integers $k_1, \ldots, k_m$.

The case $\mathbb{F} = \mathbb{C}$ follows immediately from 5.27(b) and the generalized eigenspace decomposition (8.22); thus this exercise is interesting only when $\mathbb{F} = \mathbb{R}$.

8B.9 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that there exist $D, N \in \text{End}(V)$ such that $T = D + N$, the operator D is diagonalizable, N is nilpotent, and $DN = ND$.

8B.10 Suppose V is a complex inner product space, $e_1, \ldots, e_n$ is an orthonormal basis of V, and $T \in \text{End}(V)$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, each included as many times as its multiplicity. Prove that

$$|\lambda_1|^2 + \cdots + |\lambda_n|^2 \le \|Te_1\|^2 + \cdots + \|Te_n\|^2.$$

See the comment after Exercise 5 in Section 7A.

8B.11 Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial equals $(z - 7)^2(z - 8)^2$.

8B.12 Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial equals $(z - 1)(z - 5)^3$ and whose minimal polynomial equals $(z - 1)(z - 5)^2$.

8B.13 Give an example of an operator on $\mathbb{C}^4$ whose characteristic and minimal polynomials both equal $z(z - 1)^2(z - 3)$.

8B.14 Give an example of an operator on $\mathbb{C}^4$ whose characteristic polynomial equals $z(z - 1)^2(z - 3)$ and whose minimal polynomial equals $z(z - 1)(z - 3)$.

8B.15 Let T be the operator on $\mathbb{C}^4$ defined by $T(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3)$. Find the characteristic polynomial and the minimal polynomial of T.

8B.16 Let T be the operator on $\mathbb{C}^6$ defined by

$$T(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Find the characteristic polynomial and the minimal polynomial of T.

8B.17 Suppose $\mathbb{F} = \mathbb{C}$ and $P \in \text{End}(V)$ is such that $P^2 = P$. Prove that the characteristic polynomial of P is $z^m(z-1)^n$, where $m = \dim \ker P$ and $n = \dim \text{im } P$.

8B.18 Suppose $T \in \text{End}(V)$ and λ is an eigenvalue of T. Explain why the following four numbers equal each other.

(a) The exponent of $z - \lambda$ in the factorization of the minimal polynomial of T.

(b) The smallest positive integer m such that $(T - \lambda I)^m|_{G(\lambda,T)} = 0$.

(c) The smallest positive integer m such that

$$\ker(T - \lambda I)^m = \ker(T - \lambda I)^{m+1}.$$

(d) The smallest positive integer m such that

$$\text{im}(T - \lambda I)^m = \text{im}(T - \lambda I)^{m+1}.$$

8B.19 Suppose $\mathbb{F} = \mathbb{C}$ and $S \in \text{End}(V)$ is a unitary operator. Prove that the constant term in the characteristic polynomial of S has absolute value 1.

8B.20 Suppose that $\mathbb{F} = \mathbb{C}$ and $V_1, \ldots, V_m$ are nonzero subspaces of V such that

$$V = V_1 \oplus \cdots \oplus V_m.$$

Suppose $T \in \text{End}(V)$ and each V_k is invariant under T. For each k, let p_k denote the characteristic polynomial of $T|_{V_k}$. Prove that the characteristic polynomial of T equals $p_1 \cdots p_m$.

8B.21 Suppose $p, q \in \mathbb{C}[X]$ are monic polynomials with the same zeros and q is a polynomial multiple of p. Prove that there exists $T \in \text{End}(\mathbb{C}^{\deg q})$ such that the characteristic polynomial of T is q and the minimal polynomial of T is p.

This exercise implies that every monic polynomial is the characteristic polynomial of some operator.

8B.22 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_k and B_k are square matrices of the same size for each $k = 1, \dots, m$. Show that AB is a block diagonal matrix of the form

$$AB = \begin{pmatrix} A_1 B_1 & & 0 \\ & \ddots & \\ 0 & & A_m B_m \end{pmatrix}.$$

8B.23 Suppose $\mathbb{F} = \mathbb{R}$, $T \in \text{End}(V)$, and $\lambda \in \mathbb{C}$.

(a) Show that $u + iv \in G(\lambda, T_{\mathbb{C}})$ if and only if $u - iv \in G(\overline{\lambda}, T_{\mathbb{C}})$.

(b) Show that the multiplicity of λ as an eigenvalue of $T_{\mathbb{C}}$ equals the multiplicity of $\overline{\lambda}$ as an eigenvalue of $T_{\mathbb{C}}$.

(c) Use (b) and the result about the sum of the multiplicities (8.25) to show that if $\dim V$ is an odd number, then $T_{\mathbb{C}}$ has a real eigenvalue.

(d) Use (c) and the result about real eigenvalues of $T_{\mathbb{C}}$ (Exercise 17 in Section 5A) to show that if $\dim V$ is an odd number, then T has an eigenvalue (thus giving an alternative proof of 5.34).

See Exercise 33 in Section 3B for the definition of the complexification $T_{\mathbb{C}}$.

8C Consequences of Generalized Eigenspace Decomposition

Square Roots of Operators

Recall that a square root of an operator $T \in \text{End}(V)$ is an operator $R \in \text{End}(V)$ such that $R^2 = T$ (see 7.36). 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 $\mathbb{C}^3$ defined by $T(z_1, z_2, z_3) = (z_2, z_3, 0)$ does not have a square root, as you are asked to show in Exercise 1. 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.

Identity plus nilpotent has a square root
Suppose $T \in \text{End}(V)$ is nilpotent. Then $I + T$ has a square root.

Result 8.39

Proof. Consider the Taylor series for the function $\sqrt{1+x}$:

$$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots. \tag{8.40}$$

Because $a_1 = 1/2$, the formula above implies that $1 + x/2$ is a good estimate for $\sqrt{1+x}$ when x is small.

We do not find an explicit formula for the coefficients or worry about whether the infinite sum converges because we use this equation only as motivation.

Because T is nilpotent, $T^m = 0$ for some positive integer m. In 8.40, suppose we replace x with T and 1 with I. Then the infinite sum on the right side becomes a finite sum (because $T^k = 0$ for all $k \geq m$). Thus we guess that there is a square root of $I + T$ of the form

$$I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1}.$$

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 $I + T$. Now

$$\begin{aligned}
&(I + a_1 T + a_2 T^2 + a_3 T^3 + \cdots + a_{m-1} T^{m-1})^2 \\
&\quad = I + 2a_1 T + (2a_2 + a_1^2) T^2 + (2a_3 + 2a_1 a_2) T^3 + \cdots \\
&\qquad + (2a_{m-1} + \text{terms involving } a_1, \ldots, a_{m-2}) T^{m-1}.
\end{aligned}$$

We want the right side of the equation above to equal $I + T$. 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 T^3 on the right side of the equation above equals 0 (thus $a_3 = 1/16$). Continue in this fashion for each $k = 4, \ldots, m-1$, at each step solving for a_k so that the coefficient of T^k on the right side of the equation above equals 0. Actually we do not care about the explicit formula for the a_k's. We only need to know that some choice of the a_k's gives a square root of $I + T$. □

The previous lemma is valid on real and complex vector spaces. However, the result below holds only on complex vector spaces. For example, the operator of multiplication by -1 on the one-dimensional real vector space $\mathbb{R}$ has no square root.

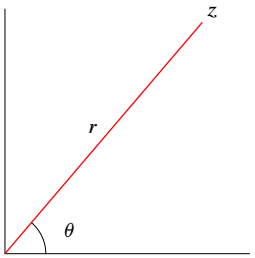
For the proof below, we need to know that every $z \in \mathbb{C}$ has a square root in $\mathbb{C}$. To show this, write

$$z = r(\cos\theta + i \sin\theta),$$

where r is the length of the line segment in the complex plane from the origin to z and θ is the angle of that line segment with the positive horizontal axis. Then

$$\sqrt{r}\left(\cos\tfrac{\theta}{2} + i \sin\tfrac{\theta}{2}\right)$$

is a square root of z, as you can verify by showing that the square of the complex number above equals z.



Representation of a complex number with polar coordinates.

Over $\mathbb{C}$, invertible operators have square roots **Result 8.41**

Suppose V is a complex vector space and $T \in \text{End}(V)$ is invertible. Then T has a square root.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. For each k, there exists a nilpotent operator $T_k \in \text{End}\big(G(\lambda_k, T)\big)$ such that $T|_{G(\lambda_k, T)} = \lambda_k I + T_k$ [see 8.22(b)]. Because T is invertible, none of the λ_k's equals 0, so we can write

$$T|_{G(\lambda_k, T)} = \lambda_k \left(I + \frac{T_k}{\lambda_k}\right)$$

for each k. Because T_k/λ_k is nilpotent, $I + T_k/\lambda_k$ has a square root (by 8.39). Multiplying a square root of the complex number λ_k by a square root of $I + T_k/\lambda_k$, we obtain a square root R_k of $T|_{G(\lambda_k,T)}$.

By the generalized eigenspace decomposition (8.22), a typical vector $v \in V$ can be written uniquely in the form

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

where each u_k is in $G(\lambda_k, T)$. Using this decomposition, define an operator $R \in \mathrm{End}(V)$ 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 subsection, you should be able to prove that if V is a complex vector space and $T \in \mathrm{End}(V)$ is invertible, then T has a kth root for every positive integer k.

Jordan Form

We know that if V is a complex vector space, then for every $T \in \mathrm{End}(V)$ there is a basis of V with respect to which T has a nice upper-triangular matrix (see 8.37). In this subsection 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.

Example 8.42 **Nilpotent operator with nice matrix**

Let T be the operator on $\mathbb{C}^4$ defined by

$$T(z_1, z_2, z_3, z_4) = (0, z_1, z_2, z_3).$$

Then $T^4 = 0$; thus T is nilpotent. If $v = (1, 0, 0, 0)$, then T^3v, T^2v, Tv, v is a basis of $\mathbb{C}^4$. The matrix of T 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.

Example 8.43 **Nilpotent operator with slightly more complicated matrix**

Example 8.43 *continued*

Let T be the operator on $\mathbb{C}^6$ defined by

$$T(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0).$$

Then $T^3 = 0$; thus T is nilpotent. In contrast to the nice behavior of the nilpotent operator of the previous example, for this nilpotent operator there does not exist a vector $v \in \mathbb{C}^6$ such that $T^5v, T^4v, T^3v, T^2v, Tv, v$ is a basis of $\mathbb{C}^6$. 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 $T^2v_1, Tv_1, v_1, Tv_2, v_2, v_3$ is a basis of $\mathbb{C}^6$. The matrix of T with respect to this basis is

$$\begin{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} & \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \begin{matrix} 0 & 0 \end{matrix} & (0) \end{pmatrix}.$$

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 goal is to show that every nilpotent operator $T \in \text{End}(V)$ behaves similarly to the operator in 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 $T^j v_k$, as k varies from 1 to n and j varies (in reverse order) from 0 to the largest nonnegative integer m_k such that $T^{m_k} v_k \neq 0$. With respect to this basis, the matrix of T looks like the matrix in the previous example. More specifically, T has a block diagonal matrix with respect to this basis, with each block a square matrix that is 0 everywhere except on the line above the diagonal.

In the next definition, the diagonal of each A_k is filled with some eigenvalue λ_k of T, the line directly above the diagonal of A_k is filled with 1's, and all other entries in A_k are 0 (to understand why each λ_k is an eigenvalue of T, see 5.41). The λ_k's need not be distinct. Also, A_k may be a 1-by-1 matrix (λ_k) containing just an eigenvalue of T. If each λ_k is 0, then the next definition captures the behavior described in the paragraph above (recall that if T is nilpotent, then 0 is the only eigenvalue of T).

Definition 8.44

Jordan basis

Suppose $T \in \text{End}(V)$. A basis of V is called a *Jordan basis* for T if with respect to this basis T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix}$$

Definition 8.44 *continued*

in which each A_k is an upper-triangular matrix of the form

$$A_k = \begin{pmatrix} \lambda_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{pmatrix}.$$

Most of the work in proving that every operator on a finite-dimensional complex vector space has a Jordan basis occurs in proving the special case below of nilpotent operators. This special case holds on real vector spaces as well as complex vector spaces.

Every nilpotent operator has a Jordan basis Result 8.45

Suppose $T \in \text{End}(V)$ is nilpotent. Then there is a basis of V that is a Jordan basis for T.

Proof. We will prove this result by induction on $\dim V$. To get started, note that the desired result holds if $\dim V = 1$ (because in that case, the only nilpotent operator is the 0 operator). Now assume that $\dim V > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Let m be the smallest positive integer such that $T^m = 0$. Thus there exists $u \in V$ such that $T^{m-1}u \neq 0$. Let

$$U = \text{span}(u, Tu, \dots, T^{m-1}u).$$

The list $u, Tu, \dots, T^{m-1}u$ is linearly independent (see Exercise 2 in Section 8A). If $U = V$, then writing this list in reverse order gives a Jordan basis for T and we are done. Thus we can assume that $U \neq V$.

Note that U is invariant under T. By our induction hypothesis, there is a basis of U that is a Jordan basis for $T|_U$. The strategy of our proof is that we will find a subspace W of V such that W is also invariant under T and $V = U \oplus W$. Again by our induction hypothesis, there will be a basis of W that is a Jordan basis for $T|_W$. Putting together the Jordan bases for $T|_U$ and $T|_W$, we will have a Jordan basis for T.

Let $\varphi \in V'$ be such that $\varphi(T^{m-1}u) \neq 0$. Let

$$W = \{v \in V : \varphi(T^k v) = 0 \text{ for each } k = 0, \dots, m-1\}.$$

Then W is a subspace of V that is invariant under T (the invariance holds because if $v \in W$ then $\varphi\big(T^k(Tv)\big) = 0$ for $k = 0, \dots, m-1$, where the case $k = m-1$ holds because $T^m = 0$). We will show that $V = U \oplus W$, which by the previous paragraph will complete the proof.

To show that $U + W$ is a direct sum, suppose $v \in U \cap W$ with $v \neq 0$. Because $v \in U$, there exist $c_0, \dots, c_{m-1} \in \mathbb{F}$ such that

$$v = c_0 u + c_1 Tu + \cdots + c_{m-1} T^{m-1} u.$$

Let j be the smallest index such that $c_j \neq 0$. Apply T^{m-j-1} to both sides of the equation above, getting

$$T^{m-j-1}v = c_j T^{m-1}u,$$

where we have used the equation $T^m = 0$. Now apply φ to both sides of the equation above, getting

$$\varphi(T^{m-j-1}v) = c_j \varphi(T^{m-1}u) \neq 0.$$

The equation above shows that $v \notin W$. Hence we have proved that $U \cap W = \{0\}$, which implies that $U + W$ is a direct sum (see 1.46).

To show that $U \oplus W = V$, define $S: V \to \mathbb{F}^m$ by

$$Sv = \big(\varphi(v), \varphi(Tv), \ldots, \varphi(T^{m-1}v)\big).$$

Thus $\ker S = W$. Hence

$$\dim W = \dim \ker S = \dim V - \dim \operatorname{im} S \geq \dim V - m,$$

where the second equality comes from the fundamental theorem of linear maps (3.21). Using the inequality above, we have

$$\dim(U \oplus W) = \dim U + \dim W \geq m + (\dim V - m) = \dim V.$$

Thus $U \oplus W = V$ (by 2.39), completing the proof. □

Camille Jordan (1838–1922) published a proof of 8.46 in 1870.

Now the generalized eigenspace decomposition allows us to extend the previous result to operators that may not be nilpotent. Doing this requires that we deal with complex vector spaces.

Result 8.46

Jordan form
Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \operatorname{End}(V)$. Then there is a basis of V that is a Jordan basis for T.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. The generalized eigenspace decomposition states that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent (see 8.22). Thus 8.45 implies that some basis of each $G(\lambda_k, T)$ is a Jordan basis for $(T - \lambda_k I)|_{G(\lambda_k, T)}$. Put these bases together to get a basis of V that is a Jordan basis for T. □

Exercises

8C.1 Suppose $T \in \text{End}(\mathbb{C}^3)$ is the operator defined by $T(z_1, z_2, z_3) = (z_2, z_3, 0)$. Prove that T does not have a square root.

8C.2 Define $T \in \text{End}(\mathbb{F}^5)$ by $T(x_1, x_2, x_3, x_4, x_5) = (2x_2, 3x_3, -x_4, 4x_5, 0)$.
(a) Show that T is nilpotent.
(b) Find a square root of $I + T$.

8C.3 Suppose V is a complex vector space. Prove that every invertible operator on V has a cube root.

8C.4 Suppose V is a real vector space. Prove that the operator $-I$ on V has a square root if and only if $\dim V$ is an even number.

8C.5 Suppose $T \in \text{End}(\mathbb{C}^2)$ is the operator defined by $T(w, z) = (-w - z, 9w + 5z)$. Find a Jordan basis for T.

8C.6 Find a basis of $\mathbb{R}[X]_{\leq 4}$ that is a Jordan basis for the differentiation operator D on $\mathbb{R}[X]_{\leq 4}$ defined by $Dp = p'$.

8C.7 Suppose $T \in \text{End}(V)$ is nilpotent and $v_1, \ldots, v_n$ is a Jordan basis for T. Prove that the minimal polynomial of T 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 T with respect to $v_1, \ldots, v_n$.

8C.8 Suppose $T \in \text{End}(V)$ 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.

8C.9 Suppose $T \in \text{End}(V)$ is nilpotent. Explain why there exist $v_1, \ldots, v_n \in V$ and nonnegative integers $m_1, \ldots, m_n$ such that (a) and (b) below both hold.
(a) $T^{m_1} v_1, \ldots, T v_1, v_1, \ldots, T^{m_n} v_n, \ldots, T v_n, v_n$ is a basis of V.
(b) $T^{m_1+1} v_1 = \cdots = T^{m_n+1} v_n = 0$.

8C.10 Suppose $T \in \text{End}(V)$ 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.

8C.11 Suppose $T \in \text{End}(V)$. Explain why every vector in each Jordan basis for T is a generalized eigenvector of T.

8C.12 Suppose $T \in \text{End}(V)$ is diagonalizable. Show that $\text{Mat}(T)$ is a diagonal matrix with respect to every Jordan basis for T.

8C.13 Suppose $T \in \text{End}(V)$ is nilpotent. Prove that if $v_1, \dots, v_n$ are vectors in V and $m_1, \dots, m_n$ are nonnegative integers such that

$$T^{m_1} v_1, \dots, T v_1, v_1, \dots, T^{m_n} v_n, \dots, T v_n, v_n \text{ is a basis of } V$$

and

$$T^{m_1+1} v_1 = \cdots = T^{m_n+1} v_n = 0,$$

then $T^{m_1} v_1, \dots, T^{m_n} v_n$ is a basis of $\ker T$.

This exercise shows that $n = \dim \ker T$. Thus the positive integer n that appears above depends only on T and not on the specific Jordan basis chosen for T.

8C.14 Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Prove that there does not exist a direct sum decomposition of V into two nonzero subspaces invariant under T if and only if the minimal polynomial of T is of the form $(z - \lambda)^{\dim V}$ for some $\lambda \in \mathbb{C}$.

8D Trace: A Connection Between Matrices and Operators

We begin this section by defining the trace of a square matrix. After developing some properties of the trace of a square matrix, we will use this concept to define the trace of an operator.

Trace of a matrix — Definition 8.47

Suppose A is a square matrix with entries in $\mathbb{F}$. The *trace* of A, denoted $\text{tr } A$, is defined to be the sum of the diagonal entries of A.

Example 8.48 **Trace of a 3-by-3 matrix**

Suppose

$$A = \begin{pmatrix} 3 & -1 & -2 \\ 5 & 2 & -3 \\ 1 & 6 & 0 \end{pmatrix}.$$

The diagonal entries of A, which are shown in red above, are 3, 2, and 0. Thus $\text{tr } A = 3 + 2 + 0 = 5$.

Matrix multiplication is not commutative, but the next result shows that the order of matrix multiplication does not matter to the trace.

Trace of AB equals trace of BA **Result 8.49**

Suppose A is an m-by-n matrix and B is an n-by-m matrix. Then

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Proof. Suppose

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}, \quad B = \begin{pmatrix} B_{1,1} & \cdots & B_{1,m} \\ \vdots & & \vdots \\ B_{n,1} & \cdots & B_{n,m} \end{pmatrix}.$$

The jth term on the diagonal of the m-by-m matrix AB equals $\sum_{k=1}^{n} A_{j,k} B_{k,j}$. Thus

$$\begin{aligned} \operatorname{tr}(AB) &= \sum_{j=1}^{m} \sum_{k=1}^{n} A_{j,k} B_{k,j} \\ &= \sum_{k=1}^{n} \sum_{j=1}^{m} B_{k,j} A_{j,k} \\ &= \sum_{k=1}^{n} (k^{\text{th}} \text{ term on diagonal of the } n\text{-by-}n \text{ matrix } BA) \\ &= \operatorname{tr}(BA), \end{aligned}$$

as desired. □

We want to define the trace of an operator $T \in \operatorname{End}(V)$ to be the trace of the matrix of T with respect to some basis of V. However, this definition should not depend on the choice of basis. The following result will make this possible.

Trace of matrix of operator does not depend on basis **Result 8.50**

Suppose $T \in \operatorname{End}(V)$. Suppose $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$ are bases of V. Then

$$\operatorname{tr} \operatorname{Mat}\big(T, (u_1, \ldots, u_n)\big) = \operatorname{tr} \operatorname{Mat}\big(T, (v_1, \ldots, v_n)\big).$$

Proof. Let $A = \operatorname{Mat}\big(T, (u_1, \ldots, u_n)\big)$ and $B = \operatorname{Mat}\big(T, (v_1, \ldots, v_n)\big)$. The change-of-basis formula tells us that there exists an invertible n-by-n matrix C such that $A = C^{-1}BC$ (see 3.84). Thus

$$\begin{aligned} \operatorname{tr} A &= \operatorname{tr}\big((C^{-1}B)C\big) \\ &= \operatorname{tr}\big(C(C^{-1}B)\big) \\ &= \operatorname{tr}\big((CC^{-1})B\big) \\ &= \operatorname{tr} B, \end{aligned}$$

where the second line comes from 8.49. □

Because of 8.50, the following definition now makes sense.

Trace of an operator — Definition 8.51

Suppose $T \in \text{End}(V)$. The *trace* of T, denoted $\text{tr}\, T$, is defined by

$$\text{tr}\, T = \text{tr}\, \text{Mat}\big(T, (v_1, \ldots, v_n)\big),$$

where $v_1, \ldots, v_n$ is any basis of V.

Suppose $T \in \text{End}(V)$ and λ is an eigenvalue of T. Recall that we defined the multiplicity of λ to be the dimension of the generalized eigenspace $G(\lambda, T)$ (see 8.23); we proved that this multiplicity equals $\dim \ker(T - \lambda I)^{\dim V}$ (see 8.20). Recall also that if V is a complex vector space, then the sum of the multiplicities of all eigenvalues of T equals $\dim V$ (see 8.25).

In the following result, the sum of the eigenvalues "with each eigenvalue included as many times as its multiplicity" means that if $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T with multiplicities $d_1, \ldots, d_m$, then the sum is

$$d_1\lambda_1 + \cdots + d_m\lambda_m.$$

Or if you prefer to work with a list of not-necessarily-distinct eigenvalues, with each eigenvalue included as many times as its multiplicity, then the eigenvalues could be denoted by $\lambda_1, \ldots, \lambda_n$ (where n equals $\dim V$) and the sum is

$$\lambda_1 + \cdots + \lambda_n.$$

On complex vector spaces, trace equals sum of eigenvalues — Result 8.52

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Then $\text{tr}\, T$ equals the sum of the eigenvalues of T, with each eigenvalue included as many times as its multiplicity.

Proof. There is a basis of V with respect to which T has an upper-triangular matrix with the diagonal entries of the matrix consisting of the eigenvalues of T, with each eigenvalue included as many times as its multiplicity—see 8.37. Thus the definition of the trace of an operator along with 8.50, which allows us to use a basis of our choice, implies that $\text{tr}\, T$ equals the sum of the eigenvalues of T, with each eigenvalue included as many times as its multiplicity. □

Example 8.53 **Trace of an operator on $\mathbb{C}^3$**

Suppose $T \in \text{End}(\mathbb{C}^3)$ is defined by

$$T(z_1, z_2, z_3) = (3z_1 - z_2 - 2z_3,\ 3z_1 + 2z_2 - 3z_3,\ z_1 + 2z_2).$$

Then the matrix of T with respect to the standard basis of $\mathbb{C}^3$ is

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

Example 8.53
continued

Adding up the diagonal entries of this matrix, we see that $\operatorname{tr} T = 5$.

The eigenvalues of T are 1, $2 + 3i$, and $2 - 3i$, each with multiplicity 1, as you can verify. The sum of these eigenvalues, each included as many times as its multiplicity, is $1 + (2 + 3i) + (2 - 3i)$, which equals 5, as expected by 8.52.

The trace has a close connection with the characteristic polynomial. Suppose $\mathbb{F} = \mathbb{C}$, $T \in \operatorname{End}(V)$, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T, with each eigenvalue included as many times as its multiplicity. Then by definition (see 8.26), 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

$$z^n - (\lambda_1 + \cdots + \lambda_n) z^{n-1} + \cdots + (-1)^n (\lambda_1 \cdots \lambda_n).$$

The expression above immediately leads to the next result. Also see 9.65, which does not require the hypothesis that $\mathbb{F} = \mathbb{C}$.

Trace and characteristic polynomial — Result 8.54

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \operatorname{End}(V)$. Let $n = \dim V$. Then $\operatorname{tr} T$ equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T.

The next result gives a nice formula for the trace of an operator on an inner product space.

Trace on an inner product space — Result 8.55

Suppose V is an inner product space, $T \in \operatorname{End}(V)$, and $e_1, \ldots, e_n$ is an orthonormal basis of V. Then

$$\operatorname{tr} T = \langle Te_1, e_1 \rangle + \cdots + \langle Te_n, e_n \rangle.$$

Proof. The desired formula follows from the observation that the entry in row k, column k of $\operatorname{Mat}\big(T, (e_1, \ldots, e_n)\big)$ equals $\langle Te_k, e_k \rangle$ [use 6.34(a) with $v = Te_k$]. □

The algebraic properties of the trace as defined on square matrices translate to algebraic properties of the trace as defined on operators, as shown in the next result.

Trace is linear — Result 8.56

The function $\operatorname{tr}\colon \operatorname{End}(V) \to \mathbb{F}$ is a linear functional on $\operatorname{End}(V)$ such that

$$\operatorname{tr}(ST) = \operatorname{tr}(TS)$$

for all $S, T \in \operatorname{End}(V)$.

Proof. Choose a basis of V. All matrices of operators in this proof will be with respect to that basis. Suppose $S, T \in \text{End}(V)$.

If $\lambda \in \mathbb{F}$, then

$$\text{tr}(\lambda T) = \text{tr}\,\text{Mat}(\lambda T) = \text{tr}\big(\lambda\,\text{Mat}(T)\big) = \lambda\,\text{tr}\,\text{Mat}(T) = \lambda\,\text{tr}\,T,$$

where the first and last equalities come from the definition of the trace of an operator, the second equality comes from 3.38, and the third equality follows from the definition of the trace of a square matrix.

Also,

$$\text{tr}(S+T) = \text{tr}\,\text{Mat}(S+T) = \text{tr}\big(\text{Mat}(S)+\text{Mat}(T)\big) = \text{tr}\,\text{Mat}(S)+\text{tr}\,\text{Mat}(T) = \text{tr}\,S+\text{tr}\,T,$$

where the first and last equalities come from the definition of the trace of an operator, the second equality comes from 3.35, and the third equality follows from the definition of the trace of a square matrix. The two paragraphs above show that $\text{tr}\colon \text{End}(V) \to \mathbb{F}$ is a linear functional on $\text{End}(V)$.

Furthermore,

$$\begin{aligned}\text{tr}(ST) &= \text{tr}\,\text{Mat}(ST) = \text{tr}\big(\text{Mat}(S)\,\text{Mat}(T)\big)\\ &= \text{tr}\big(\text{Mat}(T)\,\text{Mat}(S)\big) = \text{tr}\,\text{Mat}(TS) = \text{tr}(TS),\end{aligned}$$

where the second and fourth equalities come from 3.43 and the crucial third equality comes from 8.49. □

The statement of the next result does not involve traces, but the short proof uses traces. When something like this happens in mathematics, then usually a good definition lurks in the background.

The equations $\text{tr}(ST) = \text{tr}(TS)$ and $\text{tr}\,I = \dim V$ uniquely characterize the trace among the linear functionals on $\text{End}(V)$—see Exercise 10.

The equation $\text{tr}(ST) = \text{tr}(TS)$ leads to our next result, which does not hold on infinite-dimensional vector spaces (see Exercise 13). However, additional hypotheses on S, T, and V lead to an infinite-dimensional generalization of the result below, with important applications to quantum theory.

Result 8.57

Identity operator is not the difference of ST and TS
There do not exist operators $S, T \in \text{End}(V)$ such that $ST - TS = I$.

Proof. Suppose $S, T \in \text{End}(V)$. Then

$$\text{tr}(ST - TS) = \text{tr}(ST) - \text{tr}(TS) = 0,$$

where both equalities come from 8.56. The trace of I equals $\dim V$, which is not 0. Because $ST - TS$ and I have different traces, they cannot be equal. □

Exercises

8D.1 Suppose V is an inner product space and $v, w \in V$. Define an operator $T \in \text{End}(V)$ by $Tu = \langle u, v\rangle w$. Find a formula for $\text{tr}\, T$.

8D.2 Suppose $P \in \text{End}(V)$ satisfies $P^2 = P$. Prove that

$$\text{tr}\, P = \dim \text{im}\, P.$$

8D.3 Suppose $T \in \text{End}(V)$ and $T^5 = T$. Prove that the real and imaginary parts of $\text{tr}\, T$ are both integers.

8D.4 Suppose V is an inner product space and $T \in \text{End}(V)$. Prove that

$$\text{tr}\, T^* = \overline{\text{tr}\, T}.$$

8D.5 Suppose V is an inner product space. Suppose $T \in \text{End}(V)$ is a positive operator and $\text{tr}\, T = 0$. Prove that $T = 0$.

8D.6 Suppose V is an inner product space and $P, Q \in \text{End}(V)$ are orthogonal projections. Prove that $\text{tr}(PQ) \geq 0$.

8D.7 Suppose $T \in \text{End}(\mathbb{C}^3)$ 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.

8D.8 Prove or give a counterexample: If $S, T \in \text{End}(V)$, then $\text{tr}(ST) = (\text{tr}\, S)(\text{tr}\, T)$.

8D.9 Suppose $T \in \text{End}(V)$ is such that $\text{tr}(ST) = 0$ for all $S \in \text{End}(V)$. Prove that $T = 0$.

8D.10 Prove that the trace is the only linear functional $\tau\colon \mathrm{End}(V) \to \mathbb{F}$ such that

$$\tau(ST) = \tau(TS)$$

for all $S, T \in \mathrm{End}(V)$ and $\tau(I) = \dim V$.

Hint: Suppose that $v_1, \dots, v_n$ is a basis of V. For $j, k \in \{1, \dots, n\}$, define $P_{j,k} \in \mathrm{End}(V)$ by $P_{j,k}(a_1 v_1 + \cdots + a_n v_n) = a_k v_j$. Prove that

$$\tau(P_{j,k}) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Then for $T \in \mathrm{End}(V)$, use the equation $T = \sum_{k=1}^{n} \sum_{j=1}^{n} \mathrm{Mat}(T)_{j,k} P_{j,k}$ to show that $\tau(T) = \mathrm{tr}\, T$.

8D.11 Suppose V and W are inner product spaces and $T \in \mathrm{Hom}(V, W)$. Prove that if $e_1, \dots, e_n$ is an orthonormal basis of V and $f_1, \dots, f_m$ is an orthonormal basis of W, then

$$\mathrm{tr}(T^*T) = \sum_{k=1}^{n} \sum_{j=1}^{m} |\langle Te_k, f_j \rangle|^2.$$

The numbers $\langle Te_k, f_j \rangle$ are the entries of the matrix of T with respect to the orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_m$. These numbers depend on the bases, but $\mathrm{tr}(T^*T)$ does not depend on a choice of bases. Thus this exercise shows that the sum of the squares of the absolute values of the matrix entries does not depend on which orthonormal bases are used.

8D.12 Suppose V and W are finite-dimensional inner product spaces.

(a) Prove that $\langle S, T \rangle = \mathrm{tr}(T^*S)$ defines an inner product on $\mathrm{Hom}(V, W)$.

(b) Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $f_1, \dots, f_m$ is an orthonormal basis of W. Show that the inner product on $\mathrm{Hom}(V, W)$ from (a) is the same as the standard inner product on $\mathbb{F}^{mn}$, where we identify each element of $\mathrm{Hom}(V, W)$ with its matrix (with respect to the bases just mentioned) and then with an element of $\mathbb{F}^{mn}$.

*Caution: The norm of a linear map $T \in \mathrm{Hom}(V, W)$ as defined by 7.86 is not the same as the norm that comes from the inner product in (a) above. Unless explicitly stated otherwise, always assume that $\|T\|$ refers to the norm as defined by 7.86. The norm that comes from the inner product in (a) is called the **Frobenius norm** or the **Hilbert–Schmidt norm**.*

8D.13 Find $S, T \in \mathrm{End}\big(\mathbb{F}[X]\big)$ such that $ST - TS = I$.

Hint: Make an appropriate modification of the operators in Example 3.9.

This exercise shows that additional hypotheses are needed on S and T to extend 8.57 to the setting of infinite-dimensional vector spaces.

Multilinear Algebra and Determinants



We begin this chapter by investigating bilinear forms and quadratic forms on a vector space. Then we will move on to multilinear forms. We will show that the vector space of alternating n-linear forms has dimension one on a vector space of dimension n. This result will allow us to give a clean basis-free definition of the determinant of an operator.

This approach to the determinant via alternating multilinear forms leads to straightforward proofs of key properties of determinants. For example, we will see that the determinant is multiplicative, meaning that $\det(ST) = (\det S)(\det T)$ for all operators S and T on the same vector space. We will also see that T is invertible if and only if $\det T \neq 0$. Another important result states that the determinant of an operator on a complex vector space equals the product of the eigenvalues of the operator, with each eigenvalue included as many times as its multiplicity.

The chapter concludes with an introduction to tensor products.

Standing assumptions for this chapter — Definition

- $\mathbb{F}$ denotes $\mathbb{R}$ or $\mathbb{C}$.
- V and W denote finite-dimensional nonzero vector spaces over $\mathbb{F}$.

9A Bilinear Forms and Quadratic Forms

Bilinear Forms

A bilinear form on V is a function from $V \times V$ to $\mathbb{F}$ that is linear in each slot separately, meaning that if we hold either slot fixed then we have a linear function in the other slot. Here is the formal definition.

Bilinear form — Definition 9.1

A *bilinear form* on V is a function $\beta : V \times V \to \mathbb{F}$ such that

$$v \mapsto \beta(v, u) \quad \text{and} \quad v \mapsto \beta(u, v)$$

Definition 9.1 *continued*

are both linear functionals on V for every $u \in V$.

For example, if V is a real inner product space, then the function that takes an ordered pair $(u, v) \in V \times V$ to $\langle u, v \rangle$ is a bilinear form on V. If V is a nonzero complex inner product space, then this function is not a bilinear form because the inner product is not linear in the second slot (complex scalars come out of the second slot as their complex conjugates).

If $\mathbb{F} = \mathbb{R}$, then a bilinear form differs from an inner product in that an inner product requires symmetry [meaning that $\beta(v, w) = \beta(w, v)$ for all $v, w \in V$] and positive definiteness [meaning that $\beta(v, v) > 0$ for all $v \in V \setminus \{0\}$], but these properties are not required for a bilinear form.

Recall that the term **linear functional**, used in the definition above, means a linear function that maps into the scalar field $\mathbb{F}$. Thus the term **bilinear functional** would be more consistent terminology than **bilinear form**, which unfortunately has become standard.

Example 9.2 **Bilinear forms**

- The function $\beta: \mathbb{F}^3 \times \mathbb{F}^3 \to \mathbb{F}$ defined by
$$\beta\big((x_1, x_2, x_3), (y_1, y_2, y_3)\big) = x_1 y_2 - 5 x_2 y_3 + 2 x_3 y_1$$
is a bilinear form on $\mathbb{F}^3$.
- Suppose A is an n-by-n matrix with $A_{j,k} \in \mathbb{F}$ in row j, column k. Define a bilinear form β_A on $\mathbb{F}^n$ by
$$\beta_A\big((x_1, \ldots, x_n), (y_1, \ldots, y_n)\big) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j y_k.$$
The first bullet point is a special case of this bullet point with $n = 3$ and
$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -5 \\ 2 & 0 & 0 \end{pmatrix}.$$
- Suppose V is a real inner product space and $T \in \operatorname{End}(V)$. Then the function $\beta: V \times V \to \mathbb{R}$ defined by
$$\beta(u, v) = \langle u, Tv \rangle$$
is a bilinear form on V.
- If n is a positive integer, then the function $\beta: \mathbb{R}[X]_{\le n} \times \mathbb{R}[X]_{\le n} \to \mathbb{R}$ defined by
$$\beta(p, q) = p(2) \cdot q'(3)$$
is a bilinear form on $\mathbb{R}[X]_{\le n}$.
- Suppose $\varphi, \tau \in V'$. Then the function $\beta: V \times V \to \mathbb{F}$ defined by
$$\beta(u, v) = \varphi(u) \cdot \tau(v)$$
is a bilinear form on V.

Example 9.2 *continued*

- More generally, suppose that $\varphi_1, \dots, \varphi_n, \tau_1, \dots, \tau_n \in V'$. Then the function $\beta: V \times V \to \mathbb{F}$ defined by

$$\beta(u, v) = \varphi_1(u) \cdot \tau_1(v) + \cdots + \varphi_n(u) \cdot \tau_n(v)$$

  is a bilinear form on V.

A bilinear form on V is a function from $V \times V$ to $\mathbb{F}$. Because $V \times V$ is a vector space, this raises the question of whether a bilinear form can also be a linear map from $V \times V$ to $\mathbb{F}$. Note that none of the bilinear forms in 9.2 are linear maps except in some special cases in which the bilinear form is the zero map. Exercise 3 shows that a bilinear form β on V is a linear map on $V \times V$ only if $\beta = 0$.

Definition 9.3

$V^{(2)}$
The set of bilinear forms on V is denoted by $V^{(2)}$.

With the usual operations of addition and scalar multiplication of functions, $V^{(2)}$ is a vector space.

For T an operator on an n-dimensional vector space V and a basis $e_1, \dots, e_n$ of V, we used an n-by-n matrix to provide information about T. We now do the same thing for bilinear forms on V.

Definition 9.4

Matrix of a bilinear form, $\operatorname{Mat}(\beta)$
Suppose β is a bilinear form on V and $e_1, \dots, e_n$ is a basis of V. The *matrix* of β with respect to this basis is the n-by-n matrix $\operatorname{Mat}(\beta)$ whose entry $\operatorname{Mat}(\beta)_{j,k}$ in row j, column k is given by

$$\operatorname{Mat}(\beta)_{j,k} = \beta(e_j, e_k).$$

If the basis $e_1, \dots, e_n$ is not clear from the context, then the notation $\operatorname{Mat}\big(\beta, (e_1, \dots, e_n)\big)$ is used.

Recall that $\mathbb{F}^{n,n}$ denotes the vector space of n-by-n matrices with entries in $\mathbb{F}$ and that $\dim \mathbb{F}^{n,n} = n^2$ (see 3.39 and 3.40).

Result 9.5

$\dim V^{(2)} = (\dim V)^2$
Suppose $e_1, \dots, e_n$ is a basis of V. Then the map $\beta \mapsto \operatorname{Mat}(\beta)$ is an isomorphism of $V^{(2)}$ onto $\mathbb{F}^{n,n}$. Furthermore, $\dim V^{(2)} = (\dim V)^2$.

Proof. The map $\beta \mapsto \operatorname{Mat}(\beta)$ is clearly a linear map of $V^{(2)}$ into $\mathbb{F}^{n,n}$. For $A \in \mathbb{F}^{n,n}$, define a bilinear form β_A on V by

$$\beta_A(x_1 e_1 + \cdots + x_n e_n,\ y_1 e_1 + \cdots + y_n e_n) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j y_k$$

for $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{F}$ (if $V = \mathbb{F}^n$ and $e_1, \dots, e_n$ is the standard basis of $\mathbb{F}^n$, this β_A is the same as the bilinear form β_A in the second bullet point of Example 9.2).

The linear map $\beta \mapsto \operatorname{Mat}(\beta)$ from $V^{(2)}$ to $\mathbb{F}^{n,n}$ and the linear map $A \mapsto \beta_A$ from $\mathbb{F}^{n,n}$ to $V^{(2)}$ are inverses of each other because $\beta_{\operatorname{Mat}(\beta)} = \beta$ for all $\beta \in V^{(2)}$ and $\operatorname{Mat}(\beta_A) = A$ for all $A \in \mathbb{F}^{n,n}$, as you should verify.

Thus both maps are isomorphisms and the two spaces that they connect have the same dimension. Hence $\dim V^{(2)} = \dim \mathbb{F}^{n,n} = n^2 = (\dim V)^2$. □

Recall that C^{t} denotes the transpose of a matrix C. The matrix C^{t} is obtained by interchanging the rows and the columns of C.

Composition of a bilinear form and an operator **Result 9.6**

Suppose β is a bilinear form on V and $T \in \operatorname{End}(V)$. Define bilinear forms α and ρ on V by

$$\alpha(u, v) = \beta(u, Tv) \quad \text{and} \quad \rho(u, v) = \beta(Tu, v).$$

Let $e_1, \dots, e_n$ be a basis of V. Then

$$\operatorname{Mat}(\alpha) = \operatorname{Mat}(\beta) \operatorname{Mat}(T) \quad \text{and} \quad \operatorname{Mat}(\rho) = \operatorname{Mat}(T)^{\mathrm{t}} \operatorname{Mat}(\beta).$$

Proof. If $j, k \in \{1, \dots, n\}$, then

$$\begin{aligned}
\operatorname{Mat}(\alpha)_{j,k} &= \alpha(e_j, e_k) \\
&= \beta(e_j, Te_k) \\
&= \beta\Big(e_j, \sum_{m=1}^{n} \operatorname{Mat}(T)_{m,k} e_m\Big) \\
&= \sum_{m=1}^{n} \beta(e_j, e_m) \operatorname{Mat}(T)_{m,k} \\
&= \big(\operatorname{Mat}(\beta) \operatorname{Mat}(T)\big)_{j,k}.
\end{aligned}$$

Thus $\operatorname{Mat}(\alpha) = \operatorname{Mat}(\beta) \operatorname{Mat}(T)$. The proof that $\operatorname{Mat}(\rho) = \operatorname{Mat}(T)^{\mathrm{t}} \operatorname{Mat}(\beta)$ is similar. □

The result below shows how the matrix of a bilinear form changes if we change the basis. The formula in the result below should be compared to the change-of-basis formula for the matrix of an operator (see 3.84). The two formulas are similar, except that the transpose C^{t} appears in the formula below and the inverse C^{-1} appears in the change-of-basis formula for the matrix of an operator.

Change-of-basis formula **Result 9.7**

Suppose $\beta \in V^{(2)}$. Suppose $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are bases of V. Let

$$A = \operatorname{Mat}\big(\beta, (e_1, \dots, e_n)\big) \quad \text{and} \quad B = \operatorname{Mat}\big(\beta, (f_1, \dots, f_n)\big)$$

and $C = \operatorname{Mat}\big(I, (e_1, \dots, e_n), (f_1, \dots, f_n)\big)$. Then

$$A = C^{\mathrm{t}} B C.$$

Proof. The linear map lemma (3.4) tells us that there exists an operator $T \in \text{End}(V)$ such that $Tf_k = e_k$ for each $k = 1, \dots, n$. The definition of the matrix of an operator with respect to a basis implies that

$$\text{Mat}\big(T, (f_1, \dots, f_n)\big) = C.$$

Define bilinear forms α, ρ on V by

$$\alpha(u, v) = \beta(u, Tv) \quad \text{and} \quad \rho(u, v) = \alpha(Tu, v) = \beta(Tu, Tv).$$

Then $\beta(e_j, e_k) = \beta(Tf_j, Tf_k) = \rho(f_j, f_k)$ for all $j, k \in \{1, \dots, n\}$. Thus

$$\begin{aligned} A &= \text{Mat}\big(\rho, (f_1, \dots, f_n)\big) \\ &= C^t \, \text{Mat}\big(\alpha, (f_1, \dots, f_n)\big) \\ &= C^t B C, \end{aligned}$$

where the second and third lines each follow from 9.6. □

Example 9.8 **The matrix of a bilinear form on $\mathbb{R}[X]_{\le 2}$**

Define a bilinear form β on $\mathbb{R}[X]_{\le 2}$ by $\beta(p, q) = p(2) \cdot q'(3)$. Let

$$A = \text{Mat}\big(\beta, (1, x-2, (x-3)^2)\big) \quad \text{and} \quad B = \text{Mat}\big(\beta, (1, x, x^2)\big)$$

and $C = \text{Mat}\big(I, (1, x-2, (x-3)^2), (1, x, x^2)\big)$. Then

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 & 6 \\ 0 & 2 & 12 \\ 0 & 4 & 24 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 1 & -2 & 9 \\ 0 & 1 & -6 \\ 0 & 0 & 1 \end{pmatrix}.$$

Now the change-of-basis formula 9.7 asserts that $A = C^t B C$, which you can verify with matrix multiplication using the matrices above.

Symmetric Bilinear Forms

Definition 9.9

Symmetric bilinear form, $V^{(2)}_{\text{sym}}$
A bilinear form $\rho \in V^{(2)}$ is called *symmetric* if

$$\rho(u, w) = \rho(w, u)$$

for all $u, w \in V$. The set of symmetric bilinear forms on V is denoted by $V^{(2)}_{\text{sym}}$.

Example 9.10 **Symmetric bilinear forms**

Example 9.10
continued

- If V is a real inner product space and $\rho \in V^{(2)}$ is defined by
$$\rho(u, w) = \langle u, w \rangle,$$
then ρ is a symmetric bilinear form on V.
- Suppose V is a real inner product space and $T \in \text{End}(V)$. Define $\rho \in V^{(2)}$ by
$$\rho(u, w) = \langle u, Tw \rangle.$$
Then ρ is a symmetric bilinear form on V if and only if T is a self-adjoint operator (the previous bullet point is the special case $T = I$).
- Suppose $\rho\colon \text{End}(V) \times \text{End}(V) \to \mathbb{F}$ is defined by
$$\rho(S, T) = \text{tr}(ST).$$
Then ρ is a symmetric bilinear form on $\text{End}(V)$ because trace is a linear functional on $\text{End}(V)$ and $\text{tr}(ST) = \text{tr}(TS)$ for all $S, T \in \text{End}(V)$; see 8.56.

Symmetric matrix Definition 9.11

A square matrix A is called *symmetric* if it equals its transpose.

An operator on V may have a symmetric matrix with respect to some but not all bases of V. In contrast, the next result shows that a bilinear form on V has a symmetric matrix with respect to either all bases of V or with respect to no bases of V.

Symmetric bilinear forms are diagonalizable Result 9.12

Suppose $\rho \in V^{(2)}$. Then the following are equivalent.

(a) ρ is a symmetric bilinear form on V.

(b) $\text{Mat}\big(\rho, (e_1, \dots, e_n)\big)$ is a symmetric matrix for every basis $e_1, \dots, e_n$ of V.

(c) $\text{Mat}\big(\rho, (e_1, \dots, e_n)\big)$ is a symmetric matrix for some basis $e_1, \dots, e_n$ of V.

(d) $\text{Mat}\big(\rho, (e_1, \dots, e_n)\big)$ is a diagonal matrix for some basis $e_1, \dots, e_n$ of V.

Proof. First suppose (a) holds, so ρ is a symmetric bilinear form. Suppose $e_1, \dots, e_n$ is a basis of V and $j, k \in \{1, \dots, n\}$. Then $\rho(e_j, e_k) = \rho(e_k, e_j)$ because ρ is symmetric. Thus $\text{Mat}\big(\rho, (e_1, \dots, e_n)\big)$ is a symmetric matrix, showing that (a) implies (b).

Clearly (b) implies (c).

Now suppose (c) holds and $e_1, \dots, e_n$ is a basis of V such that $\text{Mat}\big(\rho, (e_1, \dots, e_n)\big)$ is a symmetric matrix. Suppose $u, w \in V$. There exist $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{F}$ such that $u = a_1 e_1 + \cdots + a_n e_n$ and $w = b_1 e_1 + \cdots + b_n e_n$. Now

$$\rho(u, w) = \rho\Big(\sum_{j=1}^{n} a_j e_j, \sum_{k=1}^{n} b_k e_k\Big)$$

$$\begin{aligned}
&= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j b_k \rho(e_j, e_k) \\
&= \sum_{j=1}^{n}\sum_{k=1}^{n} a_j b_k \rho(e_k, e_j) \\
&= \rho\Big(\sum_{k=1}^{n} b_k e_k, \sum_{j=1}^{n} a_j e_j\Big) \\
&= \rho(w, u),
\end{aligned}$$

where the third line holds because $\mathrm{Mat}(\rho)$ is a symmetric matrix. The equation above shows that ρ is a symmetric bilinear form, proving that (c) implies (a).

At this point, we have proved that (a), (b), (c) are equivalent. Because every diagonal matrix is symmetric, (d) implies (c). To complete the proof, we will show that (a) implies (d) by induction on $n = \dim V$.

If $n = 1$, then (a) implies (d) because every 1-by-1 matrix is diagonal. Now suppose $n > 1$ and the implication (a) $\implies$ (d) holds for one less dimension. Suppose (a) holds, so ρ is a symmetric bilinear form. If $\rho = 0$, then the matrix of ρ with respect to every basis of V is the zero matrix, which is a diagonal matrix. Hence we can assume that $\rho \neq 0$, which means there exist $u, w \in V$ such that $\rho(u, w) \neq 0$. Now

$$2\rho(u, w) = \rho(u + w, u + w) - \rho(u, u) - \rho(w, w).$$

Because the left side of the equation above is nonzero, the three terms on the right cannot all equal 0. Hence there exists $v \in V$ such that $\rho(v, v) \neq 0$.

Let $U = \{u \in V : \rho(u, v) = 0\}$. Thus U is the null space of the linear functional $u \mapsto \rho(u, v)$ on V. This linear functional is not the zero linear functional because $v \notin U$. Thus $\dim U = n - 1$. By our induction hypothesis, there is a basis $e_1, \ldots, e_{n-1}$ of U such that the symmetric bilinear form $\rho|_{U \times U}$ has a diagonal matrix with respect to this basis.

Because $v \notin U$, the list $e_1, \ldots, e_{n-1}, v$ is a basis of V. Suppose $k \in \{1, \ldots, n-1\}$. Then $\rho(e_k, v) = 0$ by the construction of U. Because ρ is symmetric, we also have $\rho(v, e_k) = 0$. Thus the matrix of ρ with respect to $e_1, \ldots, e_{n-1}, v$ is a diagonal matrix, completing the proof that (a) implies (d). □

The previous result states that every symmetric bilinear form has a diagonal matrix with respect to some basis. If our vector space happens to be a real inner product space, then the next result shows that every symmetric bilinear form has a diagonal matrix with respect to some *orthonormal* basis. Note that the inner product here is unrelated to the bilinear form.

Result 9.13

Diagonalization of a symmetric bilinear form by an orthonormal basis
Suppose V is a real inner product space and ρ is a symmetric bilinear form on V. Then ρ has a diagonal matrix with respect to some orthonormal basis of V.

Proof. Let $f_1, \ldots, f_n$ be an orthonormal basis of V. Let $B = \mathrm{Mat}\big(\rho, (f_1, \ldots, f_n)\big)$. Then B is a symmetric matrix (by 9.12). Let $T \in \mathrm{End}(V)$ be the operator such that $\mathrm{Mat}\big(T, (f_1, \ldots, f_n)\big) = B$. Thus T is self-adjoint.

The real spectral theorem (7.29) states that T has a diagonal matrix with respect to some orthonormal basis $e_1, \dots, e_n$ of V. Let $C = \operatorname{Mat}\big(I, (e_1, \dots, e_n), (f_1, \dots, f_n)\big)$. Thus $C^{-1}BC$ is the matrix of T with respect to the basis $e_1, \dots, e_n$ (by 3.84). Hence $C^{-1}BC$ is a diagonal matrix. Now

$$M\big(\rho, (e_1, \dots, e_n)\big) = C^t BC = C^{-1}BC,$$

where the first equality holds by 9.7 and the second equality holds because C is a unitary matrix with real entries (which implies that $C^{-1} = C^t$; see 7.57). □

Now we turn our attention to alternating bilinear forms. Alternating multilinear forms will play a major role in our approach to determinants later in this chapter.

Alternating bilinear form, $V_{\text{alt}}^{(2)}$ Definition 9.14

A bilinear form $\alpha \in V^{(2)}$ is called *alternating* if

$$\alpha(v, v) = 0$$

for all $v \in V$. The set of alternating bilinear forms on V is denoted by $V_{\text{alt}}^{(2)}$.

Example 9.15 Alternating bilinear forms

- Suppose $n \geq 3$ and $\alpha\colon \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is defined by

$$\alpha\big((x_1, \dots, x_n), (y_1, \dots, y_n)\big) = x_1y_2 - x_2y_1 + x_1y_3 - x_3y_1.$$

  Then α is an alternating bilinear form on $\mathbb{F}^n$.
- Suppose $\varphi, \tau \in V'$. Then the bilinear form α on V defined by

$$\alpha(u, w) = \varphi(u)\tau(w) - \varphi(w)\tau(u)$$

  is alternating.

The next result shows that a bilinear form is alternating if and only if switching the order of the two inputs multiplies the output by -1.

Characterization of alternating bilinear forms Result 9.16

A bilinear form α on V is alternating if and only if

$$\alpha(u, w) = -\alpha(w, u)$$

for all $u, w \in V$.

Proof. First suppose that α is alternating. If $u, w \in V$, then

$$\begin{aligned} 0 &= \alpha(u+w, u+w) \\ &= \alpha(u,u) + \alpha(u,w) + \alpha(w,u) + \alpha(w,w) \\ &= \alpha(u,w) + \alpha(w,u). \end{aligned}$$

Thus $\alpha(u,w) = -\alpha(w,u)$, as desired.

To prove the implication in the other direction, suppose $\alpha(u,w) = -\alpha(w,u)$ for all $u, w \in V$. Then $\alpha(v,v) = -\alpha(v,v)$ for all $v \in V$, which implies that $\alpha(v,v) = 0$ for all $v \in V$. Thus α is alternating. □

Now we show that the vector space of bilinear forms on V is the direct sum of the symmetric bilinear forms on V and the alternating bilinear forms on V.

Result 9.17

$V^{(2)} = V^{(2)}_{\text{sym}} \oplus V^{(2)}_{\text{alt}}$
The sets $V^{(2)}_{\text{sym}}$ and $V^{(2)}_{\text{alt}}$ are subspaces of $V^{(2)}$. Furthermore,

$$V^{(2)} = V^{(2)}_{\text{sym}} \oplus V^{(2)}_{\text{alt}}.$$

Proof. The definition of symmetric bilinear form implies that the sum of any two symmetric bilinear forms on V is a symmetric bilinear form on V, and every scalar multiple of any symmetric bilinear form on V is a symmetric bilinear form on V. Also, the zero bilinear form is in $V^{(2)}_{\text{sym}}$. Thus $V^{(2)}_{\text{sym}}$ is a subspace of $V^{(2)}$. Similarly, the verification that $V^{(2)}_{\text{alt}}$ is a subspace of $V^{(2)}$ is straightforward.

Next, we want to show that $V^{(2)} = V^{(2)}_{\text{sym}} + V^{(2)}_{\text{alt}}$. To do this, suppose $\beta \in V^{(2)}$. Define $\rho, \alpha \in V^{(2)}$ by

$$\rho(u,w) = \frac{\beta(u,w) + \beta(w,u)}{2} \quad \text{and} \quad \alpha(u,w) = \frac{\beta(u,w) - \beta(w,u)}{2}.$$

Then $\rho \in V^{(2)}_{\text{sym}}$ and $\alpha \in V^{(2)}_{\text{alt}}$, and $\beta = \rho + \alpha$. Thus $V^{(2)} = V^{(2)}_{\text{sym}} + V^{(2)}_{\text{alt}}$.

Finally, to show that the intersection of the two subspaces under consideration equals $\{0\}$, suppose $\beta \in V^{(2)}_{\text{sym}} \cap V^{(2)}_{\text{alt}}$. If $u, w \in V$, then 9.16 implies that

$$\beta(u,w) = -\beta(w,u) = -\beta(u,w)$$

and hence $\beta(u,w) = 0$. Thus $\beta = 0$. Hence $V^{(2)} = V^{(2)}_{\text{sym}} \oplus V^{(2)}_{\text{alt}}$ (by 1.46). □

Quadratic Forms

Definition 9.18

Quadratic form associated with a bilinear form, q_β
For β a bilinear form on V, define a function $q_\beta : V \to \mathbb{F}$ by $q_\beta(v) = \beta(v,v)$. A function $q : V \to \mathbb{F}$ is called a *quadratic form* on V if there exists a bilinear form β on V such that $q = q_\beta$.

Note that if β is a bilinear form, then $q_\beta = 0$ if and only if β is alternating.

Example 9.19 **Quadratic form**

Suppose β is the bilinear form on $\mathbb{R}^3$ defined by

$$\beta\big((x_1, x_2, x_3), (y_1, y_2, y_3)\big) = x_1y_1 - 4x_1y_2 + 8x_1y_3 - 3x_3y_3.$$

Then q_β is the quadratic form on $\mathbb{R}^3$ given by the formula

$$q_\beta(x_1, x_2, x_3) = x_1^2 - 4x_1x_2 + 8x_1x_3 - 3x_3^2.$$

The quadratic form in the example above is typical of quadratic forms on $\mathbb{F}^n$, as shown in the next result.

Result 9.20 **Quadratic forms on $\mathbb{F}^n$**

Suppose n is a positive integer and q is a function from $\mathbb{F}^n$ to $\mathbb{F}$. Then q is a quadratic form on $\mathbb{F}^n$ if and only if there exist numbers $A_{j,k} \in \mathbb{F}$ for $j, k \in \{1, \dots, n\}$ such that

$$q(x_1, \dots, x_n) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j x_k$$

for all $(x_1, \dots, x_n) \in \mathbb{F}^n$.

Proof. First suppose q is a quadratic form on $\mathbb{F}^n$. Thus there exists a bilinear form β on $\mathbb{F}^n$ such that $q = q_\beta$. Let A be the matrix of β with respect to the standard basis of $\mathbb{F}^n$. Then for all $(x_1, \dots, x_n) \in \mathbb{F}^n$, we have the desired equation

$$q(x_1, \dots, x_n) = \beta\big((x_1, \dots, x_n), (x_1, \dots, x_n)\big) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j x_k.$$

Conversely, suppose there exist numbers $A_{j,k} \in \mathbb{F}$ for $j, k \in \{1, \dots, n\}$ such that

$$q(x_1, \dots, x_n) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j x_k$$

for all $(x_1, \dots, x_n) \in \mathbb{F}^n$. Define a bilinear form β on $\mathbb{F}^n$ by

$$\beta\big((x_1, \dots, x_n), (y_1, \dots, y_n)\big) = \sum_{k=1}^{n} \sum_{j=1}^{n} A_{j,k} x_j y_k.$$

Then $q = q_\beta$, as desired. □

Although quadratic forms are defined in terms of an arbitrary bilinear form, the equivalence of (a) and (b) in the result below shows that a *symmetric* bilinear form can always be used.

Result 9.21 **Characterizations of quadratic forms**

Suppose $q: V \to \mathbb{F}$ is a function. The following are equivalent.

(a) q is a quadratic form.

(b) There exists a unique symmetric bilinear form ρ on V such that $q = q_\rho$.

(c) $q(\lambda v) = \lambda^2 q(v)$ for all $\lambda \in \mathbb{F}$ and all $v \in V$, and the function

$$(u, w) \mapsto q(u + w) - q(u) - q(w)$$

is a symmetric bilinear form on V.

(d) $q(2v) = 4q(v)$ for all $v \in V$, and the function

$$(u, w) \mapsto q(u + w) - q(u) - q(w)$$

is a symmetric bilinear form on V.

Proof. First suppose (a) holds, so q is a quadratic form. Hence there exists a bilinear form β such that $q = q_\beta$. By 9.17, there exist a symmetric bilinear form ρ on V and an alternating bilinear form α on V such that $\beta = \rho + \alpha$. Now

$$q = q_\beta = q_\rho + q_\alpha = q_\rho.$$

If $\rho' \in V_{\text{sym}}^{(2)}$ also satisfies $q_{\rho'} = q$, then $q_{\rho'-\rho} = 0$; thus $\rho' - \rho \in V_{\text{sym}}^{(2)} \cap V_{\text{alt}}^{(2)}$, which implies that $\rho' = \rho$ (by 9.17). This completes the proof that (a) implies (b).

Now suppose (b) holds, so there exists a symmetric bilinear form ρ on V such that $q = q_\rho$. If $\lambda \in \mathbb{F}$ and $v \in V$ then

$$q(\lambda v) = \rho(\lambda v, \lambda v) = \lambda \rho(v, \lambda v) = \lambda^2 \rho(v, v) = \lambda^2 q(v),$$

showing that the first part of (c) holds.

If $u, w \in V$, then

$$q(u + w) - q(u) - q(w) = \rho(u + w, u + w) - \rho(u, u) - \rho(w, w) = 2\rho(u, w).$$

Thus the function $(u, w) \mapsto q(u + w) - q(u) - q(w)$ equals 2ρ, which is a symmetric bilinear form on V, completing the proof that (b) implies (c).

Clearly (c) implies (d).

Now suppose (d) holds. Let ρ be the symmetric bilinear form on V defined by

$$\rho(u, w) = \frac{q(u + w) - q(u) - q(w)}{2}.$$

If $v \in V$, then

$$\rho(v, v) = \frac{q(2v) - q(v) - q(v)}{2} = \frac{4q(v) - 2q(v)}{2} = q(v).$$

Thus $q = q_\rho$, completing the proof that (d) implies (a). □

Example 9.22 **Symmetric bilinear form associated with a quadratic form**

Suppose q is the quadratic form on $\mathbb{R}^3$ given by the formula

$$q(x_1, x_2, x_3) = x_1^2 - 4x_1x_2 + 8x_1x_3 - 3x_3^2.$$

A bilinear form β on $\mathbb{R}^3$ such that $q = q_\beta$ is given by Example 9.19, but this bilinear form is not symmetric, as promised by 9.21(b). However, the bilinear form ρ on $\mathbb{R}^3$ defined by

$$\rho\big((x_1, x_2, x_3), (y_1, y_2, y_3)\big) = x_1y_1 - 2x_1y_2 - 2x_2y_1 + 4x_1y_3 + 4x_3y_1 - 3x_3y_3$$

is symmetric and satisfies $q = q_\rho$.

The next result states that for each quadratic form we can choose a basis such that the quadratic form looks like a weighted sum of squares of the coordinates, meaning that there are no cross terms of the form x_jx_k with $j \neq k$.

Result 9.23 **Diagonalization of quadratic form**

Suppose q is a quadratic form on V.

(a) There exist a basis $e_1, \dots, e_n$ of V and $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ such that

$$q(x_1e_1 + \cdots + x_ne_n) = \lambda_1x_1^2 + \cdots + \lambda_nx_n^2$$

for all $x_1, \dots, x_n \in \mathbb{F}$.

(b) If $\mathbb{F} = \mathbb{R}$ and V is an inner product space, then the basis in (a) can be chosen to be an orthonormal basis of V.

Proof.

(a) There exists a symmetric bilinear form ρ on V such that $q = q_\rho$ (by 9.21). Now there exists a basis $e_1, \dots, e_n$ of V such that $\operatorname{Mat}\big(\rho, (e_1, \dots, e_n)\big)$ is a diagonal matrix (by 9.12). Let $\lambda_1, \dots, \lambda_n$ denote the entries on the diagonal of this matrix. Thus

$$\rho(e_j, e_k) = \begin{cases} \lambda_j & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases}$$

for all $j, k \in \{1, \dots, n\}$. If $x_1, \dots, x_n \in \mathbb{F}$, then

$$\begin{aligned} q(x_1e_1 + \cdots + x_ne_n) &= \rho(x_1e_1 + \cdots + x_ne_n,\ x_1e_1 + \cdots + x_ne_n) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{n} x_jx_k\rho(e_j, e_k) \\ &= \lambda_1x_1^2 + \cdots + \lambda_nx_n^2, \end{aligned}$$

as desired.

(b) Suppose $\mathbb{F} = \mathbb{R}$ and V is an inner product space. Then 9.13 tells us that the basis in (a) can be chosen to be an orthonormal basis of V. □

Exercises

9A.1 Prove that if β is a bilinear form on $\mathbb{F}$, then there exists $c \in \mathbb{F}$ such that

$$\beta(x, y) = cxy$$

for all $x, y \in \mathbb{F}$.

9A.2 Let $n = \dim V$. Suppose β is a bilinear form on V. Prove that there exist $\varphi_1, \ldots, \varphi_n, \tau_1, \ldots, \tau_n \in V'$ such that

$$\beta(u, v) = \varphi_1(u) \cdot \tau_1(v) + \cdots + \varphi_n(u) \cdot \tau_n(v)$$

for all $u, v \in V$.

This exercise shows that if $n = \dim V$, then every bilinear form on V is of the form given by the last bullet point of Example 9.2.

9A.3 Suppose $\beta: V \times V \to \mathbb{F}$ is a bilinear form on V and also is a linear functional on $V \times V$. Prove that $\beta = 0$.

9A.4 Suppose V is a real inner product space and β is a bilinear form on V. Show that there exists a unique operator $T \in \text{End}(V)$ such that

$$\beta(u, v) = \langle u, Tv \rangle$$

for all $u, v \in V$.

This exercise states that if V is a real inner product space, then every bilinear form on V is of the form given by the third bullet point in 9.2.

9A.5 Suppose β is a bilinear form on a real inner product space V and T is the unique operator on V such that $\beta(u, v) = \langle u, Tv \rangle$ for all $u, v \in V$ (see Exercise 4). Show that β is an inner product on V if and only if T is an invertible positive operator on V.

9A.6 Prove or give a counterexample: If ρ is a symmetric bilinear form on V, then

$$\{v \in V : \rho(v, v) = 0\}$$

is a subspace of V.

9A.7 Explain why the proof of 9.13 (diagonalization of a symmetric bilinear form by an orthonormal basis on a real inner product space) fails if the hypothesis that $\mathbb{F} = \mathbb{R}$ is dropped.

9A.8 Find formulas for $\dim V_{\text{sym}}^{(2)}$ and $\dim V_{\text{alt}}^{(2)}$ in terms of $\dim V$.

9A.9 Suppose that n is a positive integer and $V = \{p \in \mathbb{R}[X]_{\le n} : p(0) = p(1)\}$. Define $\alpha: V \times V \to \mathbb{R}$ by

$$\alpha(p,q) = \int_0^1 pq'.$$

Show that α is an alternating bilinear form on V.

9A.10 Suppose that n is a positive integer and

$$V = \{p \in \mathbb{R}[X]_{\le n} : p(0) = p(1) \text{ and } p'(0) = p'(1)\}.$$

Define $\rho: V \times V \to \mathbb{R}$ by

$$\rho(p,q) = \int_0^1 pq''.$$

Show that ρ is a symmetric bilinear form on V.

9B Alternating Multilinear Forms

Multilinear Forms

Definition 9.24

V^m

For m a positive integer, define V^m by

$$V^m = \underbrace{V \times \cdots \times V}_{m \text{ times}}.$$

Now we can define m-linear forms as a generalization of the bilinear forms that we discussed in the previous section.

Definition 9.25

m-linear form, $V^{(m)}$, multilinear form

- For m a positive integer, an *m-linear form* on V is a function $\beta: V^m \to \mathbb{F}$ that is linear in each slot when the other slots are held fixed. This means that for each $k \in \{1, \ldots, m\}$ and all $u_1, \ldots, u_m \in V$, the function

  $$v \mapsto \beta(u_1, \ldots, u_{k-1}, v, u_{k+1}, \ldots, u_m)$$

  is a linear map from V to $\mathbb{F}$.
- The set of m-linear forms on V is denoted by $V^{(m)}$.
- A function β is called a *multilinear form* on V if it is an m-linear form on V for some positive integer m.

In the definition above, the expression $\beta(u_1, \dots, u_{k-1}, v, u_{k+1}, \dots, u_m)$ means $\beta(v, u_2, \dots, u_m)$ if $k = 1$ and means $\beta(u_1, \dots, u_{m-1}, v)$ if $k = m$.

A 1-linear form on V is a linear functional on V. A 2-linear form on V is a bilinear form on V. You can verify that with the usual addition and scalar multiplication of functions, $V^{(m)}$ is a vector space.

Example 9.26 *m*-linear forms

- Suppose $\alpha, \rho \in V^{(2)}$. Define a function $\beta\colon V^4 \to \mathbb{F}$ by
$$\beta(v_1, v_2, v_3, v_4) = \alpha(v_1, v_2)\, \rho(v_3, v_4).$$
Then $\beta \in V^{(4)}$.
- Define $\beta\colon \big(\text{End}(V)\big)^m \to \mathbb{F}$ by
$$\beta(T_1, \dots, T_m) = \text{tr}(T_1 \cdots T_m).$$
Then β is an m-linear form on $\text{End}(V)$.

Alternating multilinear forms, which we now define, play an important role as we head toward defining determinants.

Definition 9.27

Alternating forms, $V_{\text{alt}}^{(m)}$

Suppose m is a positive integer.

- An m-linear form α on V is called *alternating* if $\alpha(v_1, \dots, v_m) = 0$ whenever $v_1, \dots, v_m$ is a list of vectors in V with $v_j = v_k$ for some two distinct values of j and k in $\{1, \dots, m\}$.
- $V_{\text{alt}}^{(m)} = \{\alpha \in V^{(m)} : \alpha \text{ is an alternating } m\text{-linear form on } V\}$.

You should verify that $V_{\text{alt}}^{(m)}$ is a subspace of $V^{(m)}$. See Example 9.15 for examples of alternating 2-linear forms. See Exercise 2 for an example of an alternating 3-linear form.

The next result tells us that if a linearly dependent list is input to an alternating multilinear form, then the output equals 0.

Result 9.28

Alternating multilinear forms and linear dependence

Suppose m is a positive integer and α is an alternating m-linear form on V. If $v_1, \dots, v_m$ is a linearly dependent list in V, then
$$\alpha(v_1, \dots, v_m) = 0.$$

Proof. Suppose $v_1, \dots, v_m$ is a linearly dependent list in V. By the linear dependence lemma (2.19), some v_k is a linear combination of $v_1, \dots, v_{k-1}$. Thus there

exist $b_1, \dots, b_{k-1}$ such that $v_k = b_1 v_1 + \cdots + b_{k-1} v_{k-1}$. Now

$$\begin{aligned}\alpha(v_1, \dots, v_m) &= \alpha\Big(v_1, \dots, v_{k-1}, \sum_{j=1}^{k-1} b_j v_j, v_{k+1}, \dots, v_m\Big) \\ &= \sum_{j=1}^{k-1} b_j \, \alpha(v_1, \dots, v_{k-1}, v_j, v_{k+1}, \dots, v_m) \\ &= 0.\end{aligned}$$

□

The next result states that if $m > \dim V$, then there are no alternating m-linear forms on V other than the function on V^m that is identically 0.

No nonzero alternating m-linear forms for $m > \dim V$ **Result 9.29**
Suppose $m > \dim V$. Then 0 is the only alternating m-linear form on V.

Proof. Suppose that α is an alternating m-linear form on V and $v_1, \dots, v_m \in V$. Because $m > \dim V$, this list is not linearly independent (by 2.22). Thus 9.28 implies that $\alpha(v_1, \dots, v_m) = 0$. Hence α is the zero function from V^m to $\mathbb{F}$. □

Alternating Multilinear Forms and Permutations

Swapping input vectors in an alternating multilinear form **Result 9.30**
Suppose m is a positive integer, α is an alternating m-linear form on V, and $v_1, \dots, v_m$ is a list of vectors in V. Then swapping the vectors in any two slots of $\alpha(v_1, \dots, v_m)$ changes the value of α by a factor of -1.

Proof. Put $v_1 + v_2$ in both the first two slots, getting

$$0 = \alpha(v_1 + v_2, v_1 + v_2, v_3, \dots, v_m).$$

Use the multilinear properties of α to expand the right side of the equation above (as in the proof of 9.16) to get

$$\alpha(v_2, v_1, v_3, \dots, v_m) = -\alpha(v_1, v_2, v_3, \dots, v_m).$$

Similarly, swapping the vectors in any two slots of $\alpha(v_1, \dots, v_m)$ changes the value of α by a factor of -1. □

To see what can happen with multiple swaps, suppose α is an alternating 3-linear form on V and $v_1, v_2, v_3 \in V$. To evaluate $\alpha(v_3, v_1, v_2)$ in terms of $\alpha(v_1, v_2, v_3)$, start with $\alpha(v_3, v_1, v_2)$ and swap the entries in the first and third slots, getting $\alpha(v_3, v_1, v_2) = -\alpha(v_2, v_1, v_3)$. Now in the last expression, swap the entries in the first and second slots, getting

$$\alpha(v_3, v_1, v_2) = -\alpha(v_2, v_1, v_3) = \alpha(v_1, v_2, v_3).$$

More generally, we see that if we do an odd number of swaps, then the value of α changes by a factor of -1, and if we do an even number of swaps, then the value of α does not change.

To deal with arbitrary multiple swaps, we need a bit of information about permutations.

Permutation, perm m — Definition 9.31

Suppose m is a positive integer.

- A *permutation* of $(1, \dots, m)$ is a list $(j_1, \dots, j_m)$ that contains each of the numbers $1, \dots, m$ exactly once.
- The set of all permutations of $(1, \dots, m)$ is denoted by perm m.

For example, $(2, 3, 4, 5, 1) \in \operatorname{perm} 5$. You should think of an element of perm m as a rearrangement of the first m positive integers.

The number of swaps used to change a permutation $(j_1, \dots, j_m)$ to the standard order $(1, \dots, m)$ can depend on the specific swaps selected. The following definition has the advantage of assigning a well-defined sign to every permutation.

Sign of a permutation — Definition 9.32

The *sign* of a permutation $(j_1, \dots, j_m)$ is defined by

$$\operatorname{sign}(j_1, \dots, j_m) = (-1)^N,$$

where N is the number of pairs of integers (k, ℓ) with $1 \le k < \ell \le m$ such that k appears after ℓ in the list $(j_1, \dots, j_m)$.

Hence 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.

Example 9.33 **Signs**

- The permutation $(1, \dots, m)$ [no changes in the natural order] has sign 1.
- The only pair of integers (k, ℓ) with $k < \ell$ such that k appears after ℓ in the list $(2, 1, 3, 4)$ is $(1, 2)$. Thus the permutation $(2, 1, 3, 4)$ has sign -1.
- In the permutation $(2, 3, \dots, m, 1)$, the only pairs (k, ℓ) with $k < \ell$ that appear with changed order are $(1, 2), (1, 3), \dots, (1, m)$. Because we have $m - 1$ such pairs, the sign of this permutation equals $(-1)^{m-1}$.

Swapping two entries in a permutation **Result 9.34**

Swapping 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 swapping two entries. The two swapped entries were in their natural order in the first permutation if and only if they are not in their natural order in the second permutation. Thus we have a net change (so far) of 1 or -1 (both odd numbers) in the number of pairs not in their natural order.

Consider each entry between the two swapped entries. If an intermediate entry was originally in the natural order with respect to both swapped entries, then it is now in the natural order with respect to neither swapped entry. Similarly, if an intermediate entry was originally in the natural order with respect to neither of the swapped entries, then it is now in the natural order with respect to both swapped entries. If an intermediate entry was originally in the natural order with respect to exactly one of the swapped entries, then that is still true. Thus the net change (for each pair containing an entry between the two swapped entries) in the number of pairs not in their natural order is 2, -2, or 0 (all even numbers).

For all other pairs of entries, there is no change in whether or not they are in their natural order. Thus the total net change in the number of pairs not in their natural order is an odd number. Hence the sign of the second permutation equals -1 times the sign of the first permutation. □

Permutations and alternating multilinear forms **Result 9.35**

Suppose m is a positive integer and $\alpha \in V_{\text{alt}}^{(m)}$. Then

$$\alpha(v_{j_1}, \ldots, v_{j_m}) = \big(\text{sign}(j_1, \ldots, j_m)\big)\alpha(v_1, \ldots, v_m)$$

for every list $v_1, \ldots, v_m$ of vectors in V and all $(j_1, \ldots, j_m) \in \text{perm}\, m$.

Proof. Suppose $v_1, \ldots, v_m \in V$ and $(j_1, \ldots, j_m) \in \text{perm}\, m$. We can get from $(j_1, \ldots, j_m)$ to $(1, \ldots, m)$ by a series of swaps of entries in different slots. Each such swap changes the value of α by a factor of -1 (by 9.30) and also changes the sign of the remaining permutation by a factor of -1 (by 9.34). After an appropriate number of swaps, we reach the permutation $1, \ldots, m$, which has sign 1. Thus the value of α changed signs an even number of times if $\text{sign}(j_1, \ldots, j_m) = 1$ and an odd number of times if $\text{sign}(j_1, \ldots, j_m) = -1$, which gives the desired result. □

Our use of permutations now leads in a natural way to the following beautiful formula for alternating n-linear forms on an n-dimensional vector space.

Formula for $(\dim V)$-linear alternating forms on V **Result 9.36**

Let $n = \dim V$. Suppose $e_1, \ldots, e_n$ is a basis of V and $v_1, \ldots, v_n \in V$. For each

Result 9.36 *continued*

$k \in \{1, \dots, n\}$, let $b_{1,k}, \dots, b_{n,k} \in \mathbb{F}$ be such that

$$v_k = \sum_{j=1}^{n} b_{j,k} e_j.$$

Then

$$\alpha(v_1, \dots, v_n) = \alpha(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \dots, j_n)\big) b_{j_1,1} \cdots b_{j_n,n}$$

for every alternating n-linear form α on V.

Proof. Suppose α is an alternating n-linear form on V. Then

$$\begin{aligned}
\alpha(v_1, \dots, v_n) &= \alpha\Big(\sum_{j_1=1}^{n} b_{j_1,1} e_{j_1}, \dots, \sum_{j_n=1}^{n} b_{j_n,n} e_{j_n}\Big) \\
&= \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} b_{j_1,1} \cdots b_{j_n,n}\, \alpha(e_{j_1}, \dots, e_{j_n}) \\
&= \sum_{(j_1,\dots,j_n) \in \operatorname{perm} n} b_{j_1,1} \cdots b_{j_n,n}\, \alpha(e_{j_1}, \dots, e_{j_n}) \\
&= \alpha(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \dots, j_n)\big) b_{j_1,1} \cdots b_{j_n,n},
\end{aligned}$$

where the third line holds because $\alpha(e_{j_1}, \dots, e_{j_n}) = 0$ if $j_1, \dots, j_n$ are not distinct integers, and the last line holds by 9.35. □

The following result will be the key to our definition of the determinant in the next section.

Result 9.37

$\dim V_{\text{alt}}^{(\dim V)} = 1$
The vector space $V_{\text{alt}}^{(\dim V)}$ has dimension one.

Proof. Let $n = \dim V$. Suppose α and α' are alternating n-linear forms on V with $\alpha \neq 0$. Let $e_1, \dots, e_n$ be such that $\alpha(e_1, \dots, e_n) \neq 0$. There exists $c \in \mathbb{F}$ such that

$$\alpha'(e_1, \dots, e_n) = c\alpha(e_1, \dots, e_n).$$

Furthermore, 9.28 implies that $e_1, \dots, e_n$ is linearly independent and thus is a basis of V.

Suppose $v_1, \dots, v_n \in V$. Let $b_{j,k}$ be as in 9.36 for $j, k = 1, \dots, n$. Then

$$\begin{aligned}
\alpha'(v_1, \dots, v_n) &= \alpha'(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \dots, j_n)\big) b_{j_1,1} \cdots b_{j_n,n} \\
&= c\alpha(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \dots, j_n)\big) b_{j_1,1} \cdots b_{j_n,n} \\
&= c\alpha(v_1, \dots, v_n),
\end{aligned}$$

where the first and last lines above come from 9.36. The equation above implies that $\alpha' = c\alpha$. Thus α', α is not a linearly independent list, which implies that $\dim V_{\text{alt}}^{(n)} \le 1$.

To complete the proof, we only need to show that there exists a nonzero alternating n-linear form α on V (thus eliminating the possibility that $\dim V_{\text{alt}}^{(n)}$ equals 0). To do this, let $e_1, \dots, e_n$ be any basis of V, and let $\varphi_1, \dots, \varphi_n \in V'$ be the linear functionals on V that allow us to express each element of V as a linear combination of $e_1, \dots, e_n$. In other words,

$$v = \sum_{j=1}^{n} \varphi_j(v) e_j$$

for every $v \in V$ (see 3.114). Now for $v_1, \dots, v_n \in V$, define

$$\alpha(v_1, \dots, v_n) = \sum_{(j_1,\dots,j_n) \in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \dots, j_n)\big) \varphi_{j_1}(v_1) \cdots \varphi_{j_n}(v_n). \tag{9.38}$$

The verification that α is an n-linear form on V is straightforward.

To see that α is alternating, suppose $v_1, \dots, v_n \in V$ with $v_1 = v_2$. For each $(j_1, \dots, j_n) \in \operatorname{perm} n$, the permutation $(j_2, j_1, j_3, \dots, j_n)$ has the opposite sign. Because $v_1 = v_2$, the contributions from these two permutations to the sum in 9.38 cancel each other. Hence $\alpha(v_1, v_1, v_3, \dots, v_n) = 0$. Similarly, $\alpha(v_1, \dots, v_n) = 0$ if any two vectors in the list $v_1, \dots, v_n$ are equal. Thus α is alternating.

Finally, consider 9.38 with each $v_k = e_k$. Because $\varphi_j(e_k)$ equals 0 if $j \ne k$ and equals 1 if $j = k$, only the permutation $(1, \dots, n)$ makes a nonzero contribution to the right side of 9.38 in this case, giving the equation $\alpha(e_1, \dots, e_n) = 1$. Thus we have produced a nonzero alternating n-linear form α on V, as desired. □

The formula 9.38 used in the last proof to construct a nonzero alternating n-linear form came from the formula in 9.36, and that formula arose naturally from the properties of an alternating multilinear form.

Earlier we showed that the value of an alternating multilinear form applied to a linearly dependent list is 0; see 9.28. The next result provides a converse of 9.28 for alternating n-linear forms when $n = \dim V$. In the following result, the statement that α is nonzero means (as usual for a function) that α is not the function on V^n that is identically 0.

Result 9.39

Alternating $(\dim V)$-linear forms and linear independence

Let $n = \dim V$. Suppose α is a nonzero alternating n-linear form on V and $e_1, \dots, e_n$ is a list of vectors in V. Then

$$\alpha(e_1, \dots, e_n) \ne 0$$

if and only if $e_1, \dots, e_n$ is linearly independent.

Proof. First suppose $\alpha(e_1, \dots, e_n) \ne 0$. Then 9.28 implies that $e_1, \dots, e_n$ is linearly independent.

To prove the implication in the other direction, now suppose $e_1, \dots, e_n$ is linearly independent. Because $n = \dim V$, this implies that $e_1, \dots, e_n$ is a basis of V (see 2.38).

Because α is not the zero n-linear form, there exist $v_1, \dots, v_n \in V$ such that $\alpha(v_1, \dots, v_n) \ne 0$. Now 9.36 implies that $\alpha(e_1, \dots, e_n) \ne 0$. □

Exercises

9B.1 Suppose m is a positive integer. Show that $\dim V^{(m)} = (\dim V)^m$.

9B.2 Suppose $n \geq 3$ and $\alpha: \mathbb{F}^n \times \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is defined by

$$\alpha\big((x_1, \ldots, x_n), (y_1, \ldots, y_n), (z_1, \ldots, z_n)\big) \\ = x_1y_2z_3 - x_2y_1z_3 - x_3y_2z_1 - x_1y_3z_2 + x_3y_1z_2 + x_2y_3z_1.$$

Show that α is an alternating 3-linear form on $\mathbb{F}^n$.

9B.3 Suppose m is a positive integer and α is an m-linear form on V such that $\alpha(v_1, \ldots, v_m) = 0$ whenever $v_1, \ldots, v_m$ is a list of vectors in V with $v_j = v_{j+1}$ for some $j \in \{1, \ldots, m-1\}$. Prove that α is an alternating m-linear form on V.

9B.4 Prove or give a counterexample: If $\alpha \in V^{(4)}_{\text{alt}}$, then

$$\{(v_1, v_2, v_3, v_4) \in V^4 : \alpha(v_1, v_2, v_3, v_4) = 0\}$$

is a subspace of V^4.

9B.5 Suppose m is a positive integer and β is an m-linear form on V. Define an m-linear form α on V by

$$\alpha(v_1, \ldots, v_m) = \sum_{(j_1, \ldots, j_m) \in \operatorname{perm} m} \big(\operatorname{sign}(j_1, \ldots, j_m)\big)\beta(v_{j_1}, \ldots, v_{j_m})$$

for $v_1, \ldots, v_m \in V$. Explain why $\alpha \in V^{(m)}_{\text{alt}}$.

9B.6 Suppose m is a positive integer and β is an m-linear form on V. Define an m-linear form α on V by

$$\alpha(v_1, \ldots, v_m) = \sum_{(j_1, \ldots, j_m) \in \operatorname{perm} m} \beta(v_{j_1}, \ldots, v_{j_m})$$

for $v_1, \ldots, v_m \in V$. Explain why

$$\alpha(v_{k_1}, \ldots, v_{k_m}) = \alpha(v_1, \ldots, v_m)$$

for all $v_1, \ldots, v_m \in V$ and all $(k_1, \ldots, k_m) \in \operatorname{perm} m$.

9B.7 Give an example of a nonzero alternating 2-linear form α on $\mathbb{R}^3$ and a linearly independent list v_1, v_2 in $\mathbb{R}^3$ such that $\alpha(v_1, v_2) = 0$.

This exercise shows that 9.39 can fail if the hypothesis that $n = \dim V$ is deleted.

9C Determinants

Defining the Determinant

The next definition will lead us to a clean, beautiful, basis-free definition of the determinant of an operator.

Definition 9.40 α_T

Suppose that m is a positive integer and $T \in \text{End}(V)$. For $\alpha \in V_{\text{alt}}^{(m)}$, define $\alpha_T \in V_{\text{alt}}^{(m)}$ by

$$\alpha_T(v_1, \dots, v_m) = \alpha(Tv_1, \dots, Tv_m)$$

for each list $v_1, \dots, v_m$ of vectors in V.

Suppose $T \in \text{End}(V)$. If $\alpha \in V_{\text{alt}}^{(m)}$ and $v_1, \dots, v_m$ is a list of vectors in V with $v_j = v_k$ for some $j \neq k$, then $Tv_j = Tv_k$, which implies that $\alpha_T(v_1, \dots, v_m) = \alpha(Tv_1, \dots, Tv_m) = 0$. Thus the function $\alpha \mapsto \alpha_T$ is a linear map of $V_{\text{alt}}^{(m)}$ to itself.

We know that $\dim V_{\text{alt}}^{(\dim V)} = 1$ (see 9.37). Every linear map from a one-dimensional vector space to itself is multiplication by some unique scalar. For the linear map $\alpha \mapsto \alpha_T$, we now define $\det T$ to be that scalar.

Definition 9.41 **Determinant of an operator,** $\det T$

Suppose $T \in \text{End}(V)$. The *determinant* of T, denoted by $\det T$, is defined to be the unique number in $\mathbb{F}$ such that

$$\alpha_T = (\det T)\,\alpha$$

for all $\alpha \in V_{\text{alt}}^{(\dim V)}$.

Example 9.42 **Determinants of operators**

Let $n = \dim V$.

- If I is the identity operator on V, then $\alpha_I = \alpha$ for all $\alpha \in V_{\text{alt}}^{(n)}$. Thus $\det I = 1$.
- More generally, if $\lambda \in \mathbb{F}$, then $\alpha_{\lambda I} = \lambda^n \alpha$ for all $\alpha \in V_{\text{alt}}^{(n)}$. Thus $\det(\lambda I) = \lambda^n$.
- Still more generally, if $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$, then $\alpha_{\lambda T} = \lambda^n \alpha_T = \lambda^n (\det T)\alpha$ for all $\alpha \in V_{\text{alt}}^{(n)}$. Thus $\det(\lambda T) = \lambda^n \det T$.
- Suppose $T \in \text{End}(V)$ and there is a basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T, with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. If $\alpha \in V_{\text{alt}}^{(n)}$, then

  $$\alpha_T(e_1, \dots, e_n) = \alpha(\lambda_1 e_1, \dots, \lambda_n e_n) = (\lambda_1 \cdots \lambda_n)\alpha(e_1, \dots, e_n).$$

  If $\alpha \neq 0$, then 9.39 implies $\alpha(e_1, \dots, e_n) \neq 0$. Thus the equation above implies

  $$\det T = \lambda_1 \cdots \lambda_n.$$

Our next task is to define and give a formula for the determinant of a square matrix. To do this, we associate with each square matrix an operator and then define the determinant of the matrix to be the determinant of the associated operator.

Determinant of a matrix, $\det A$ — Definition 9.43

Suppose that n is a positive integer and A is an n-by-n square matrix with entries in $\mathbb{F}$. Let $T \in \text{End}(\mathbb{F}^n)$ be the operator whose matrix with respect to the standard basis of $\mathbb{F}^n$ equals A. The *determinant* of A, denoted by $\det A$, is defined by $\det A = \det T$.

Example 9.44 **Determinants of matrices**

- If I is the n-by-n identity matrix, then the corresponding operator on $\mathbb{F}^n$ is the identity operator I on $\mathbb{F}^n$. Thus the first bullet point of 9.42 implies that the determinant of the identity matrix is 1.
- Suppose A is a diagonal matrix with $\lambda_1, \dots, \lambda_n$ on the diagonal. Then the corresponding operator on $\mathbb{F}^n$ has the standard basis of $\mathbb{F}^n$ as eigenvectors, with eigenvalues $\lambda_1, \dots, \lambda_n$. Thus the last bullet point of 9.42 implies that $\det A = \lambda_1 \cdots \lambda_n$.

For the next result, think of each list $v_1, \dots, v_n$ of n vectors in $\mathbb{F}^n$ as a list of n-by-1 column vectors. The notation $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ then denotes the n-by-n square matrix whose k^{th} column is v_k for each $k = 1, \dots, n$.

Determinant is an alternating multilinear form — Result 9.45

Suppose that n is a positive integer. The map that takes a list $v_1, \dots, v_n$ of vectors in $\mathbb{F}^n$ to $\det \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ is an alternating n-linear form on $\mathbb{F}^n$.

Proof. Let $e_1, \dots, e_n$ be the standard basis of $\mathbb{F}^n$ and suppose $v_1, \dots, v_n$ is a list of vectors in $\mathbb{F}^n$. Let $T \in \text{End}(\mathbb{F}^n)$ be the operator such that $Te_k = v_k$ for $k = 1, \dots, n$. Thus T is the operator whose matrix with respect to $e_1, \dots, e_n$ is $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$. Hence $\det \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \det T$, by definition of the determinant of a matrix.

Let α be an alternating n-linear form on $\mathbb{F}^n$ such that $\alpha(e_1, \dots, e_n) = 1$. Then

$$\begin{aligned} \det \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} &= \det T \\ &= (\det T)\, \alpha(e_1, \dots, e_n) \\ &= \alpha(Te_1, \dots, Te_n) \\ &= \alpha(v_1, \dots, v_n), \end{aligned}$$

where the third line follows from the definition of the determinant of an operator. The equation above shows that the map that takes a list of vectors $v_1, \dots, v_n$ in $\mathbb{F}^n$ to $\det \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ is the alternating n-linear form α on $\mathbb{F}^n$. □

The previous result has several important consequences. For example, it immediately implies that a matrix with two identical columns has determinant 0. We will come back to other consequences later, but for now we want to give a formula for the determinant of a square matrix. Recall that if A is a matrix, then $A_{j,k}$ denotes the entry in row j, column k of A.

Formula for determinant of a matrix Result 9.46

Suppose that n is a positive integer and A is an n-by-n square matrix. Then

$$\det A = \sum_{(j_1,\dots,j_n)\in \operatorname{perm} n} \big(\operatorname{sign}(j_1,\dots,j_n)\big) A_{j_1,1}\cdots A_{j_n,n}.$$

Proof. Apply 9.36 with $V = \mathbb{F}^n$ and $e_1,\dots,e_n$ the standard basis of $\mathbb{F}^n$ and α the alternating n-linear form on $\mathbb{F}^n$ that takes $v_1,\dots,v_n$ to $\det\begin{pmatrix} v_1 & \cdots & v_n\end{pmatrix}$ [see 9.45]. If each v_k is the kth column of A, then each $b_{j,k}$ in 9.36 equals $A_{j,k}$. Finally,

$$\alpha(e_1,\dots,e_n) = \det\begin{pmatrix} e_1 & \cdots & e_n\end{pmatrix} = \det I = 1.$$

Thus the formula in 9.36 becomes the formula stated in this result. □

Example 9.47 **Explicit formula for determinant**

- If A is a 2-by-2 matrix, then the formula in 9.46 becomes
$$\det A = A_{1,1}A_{2,2} - A_{2,1}A_{1,2}.$$
- If A is a 3-by-3 matrix, then the formula in 9.46 becomes
$$\begin{aligned}\det A = {} & A_{1,1}A_{2,2}A_{3,3} - A_{2,1}A_{1,2}A_{3,3} - A_{3,1}A_{2,2}A_{1,3}\\ & - A_{1,1}A_{3,2}A_{2,3} + A_{3,1}A_{1,2}A_{2,3} + A_{2,1}A_{3,2}A_{1,3}.\end{aligned}$$

The sum in the formula in 9.46 contains $n!$ terms. Because $n!$ grows rapidly as n increases, the formula in 9.46 is not a viable method to evaluate determinants even for moderately sized n. For example, 10! is over three million, and 100! is approximately 10^{158}, leading to a sum that the fastest computer cannot evaluate. We will soon see some results that lead to faster evaluations of determinants than direct use of the sum in 9.46.

Determinant of upper-triangular matrix Result 9.48

Suppose that A is an upper-triangular matrix with $\lambda_1,\dots,\lambda_n$ on the diagonal. Then $\det A = \lambda_1\cdots\lambda_n$.

Proof. If $(j_1,\dots,j_n) \in \operatorname{perm} n$ with $(j_1,\dots,j_n) \neq (1,\dots,n)$, then $j_k > k$ for some $k \in \{1,\dots,n\}$, which implies that $A_{j_k,k} = 0$. Thus the only permutation that can

make a nonzero contribution to the sum in 9.46 is the permutation $(1, \dots, n)$. Because $A_{k,k} = \lambda_k$ for each $k = 1, \dots, n$, this implies that $\det A = \lambda_1 \cdots \lambda_n$. □

Properties of Determinants

Our definition of the determinant leads to the following magical proof that the determinant is multiplicative.

Determinant is multiplicative **Result 9.49**

(a) Suppose $S, T \in \text{End}(V)$. Then $\det(ST) = (\det S)(\det T)$.

(b) Suppose A and B are square matrices of the same size. Then

$$\det(AB) = (\det A)(\det B)$$

Proof.

(a) Let $n = \dim V$. Suppose $\alpha \in V_{\text{alt}}^{(n)}$ and $v_1, \dots, v_n \in V$. Then

$$\begin{aligned} \alpha_{ST}(v_1, \dots, v_n) &= \alpha(STv_1, \dots, STv_n) \\ &= (\det S)\alpha(Tv_1, \dots, Tv_n) \\ &= (\det S)(\det T)\alpha(v_1, \dots, v_n), \end{aligned}$$

where the first equation follows from the definition of α_{ST}, the second equation follows from the definition of $\det S$, and the third equation follows from the definition of $\det T$. The equation above implies that $\det(ST) = (\det S)(\det T)$.

(b) Let $S, T \in \text{End}(\mathbb{F}^n)$ be such that $\text{Mat}(S) = A$ and $\text{Mat}(T) = B$, where all matrices of operators in this proof are with respect to the standard basis of $\mathbb{F}^n$. Then $\text{Mat}(ST) = \text{Mat}(S)\,\text{Mat}(T) = AB$ (see 3.43). Thus

$$\det(AB) = \det(ST) = (\det S)(\det T) = (\det A)(\det B),$$

where the second equality comes from the result in (a). □

The determinant of an operator determines whether the operator is invertible.

Invertible $\iff$ nonzero determinant **Result 9.50**

An operator $T \in \text{End}(V)$ is invertible if and only if $\det T \neq 0$. Furthermore, if T is invertible, then $\det(T^{-1}) = 1/(\det T)$.

Proof. First suppose T is invertible. Thus $TT^{-1} = I$. Now 9.49 implies that

$$1 = \det I = \det(TT^{-1}) = (\det T)\big(\det(T^{-1})\big).$$

Hence $\det T \neq 0$ and $\det(T^{-1})$ is the multiplicative inverse of $\det T$.

To prove the other direction, now suppose $\det T \neq 0$. Suppose $v \in V$ and $v \neq 0$. Let $v, e_2, \ldots, e_n$ be a basis of V and let $\alpha \in V_{\text{alt}}^{(n)}$ be such that $\alpha \neq 0$. Then $\alpha(v, e_2, \ldots, e_n) \neq 0$ (by 9.39). Now

$$\alpha(Tv, Te_2, \ldots, Te_n) = (\det T)\alpha(v, e_2, \ldots, e_n) \neq 0.$$

Thus $Tv \neq 0$. Hence T is invertible. □

An n-by-n matrix A is invertible (see 3.80 for the definition of an invertible matrix) if and only if the operator on $\mathbb{F}^n$ associated with A (via the standard basis of $\mathbb{F}^n$) is invertible. Thus the previous result shows that a square matrix A is invertible if and only if $\det A \neq 0$.

Eigenvalues and determinants **Result 9.51**

Suppose $T \in \text{End}(V)$ and $\lambda \in \mathbb{F}$. Then λ is an eigenvalue of T if and only if $\det(\lambda I - T) = 0$.

Proof. The number λ is an eigenvalue of T if and only if $T - \lambda I$ is not invertible (see 5.7), which happens if and only if $\lambda I - T$ is not invertible, which happens if and only if $\det(\lambda I - T) = 0$ (by 9.50). □

Suppose $T \in \text{End}(V)$ and $S: W \to V$ is an invertible linear map. To prove that $\det(S^{-1}TS) = \det T$, we could try to use 9.49 and 9.50, writing

$$\begin{aligned}\det(S^{-1}TS) &= (\det S^{-1})(\det T)(\det S)\\ &= \det T.\end{aligned}$$

That proof works if $W = V$, but if $W \neq V$ then it makes no sense because the determinant is defined only for linear maps from a vector space to itself, and S maps W to V, making $\det S$ undefined. The proof given below works around this issue and is valid when $W \neq V$.

Determinant is a similarity invariant **Result 9.52**

Suppose $T \in \text{End}(V)$ and $S: W \to V$ is an invertible linear map. Then

$$\det(S^{-1}TS) = \det T.$$

Proof. Let $n = \dim W = \dim V$. Suppose $\tau \in W_{\text{alt}}^{(n)}$. Define $\alpha \in V_{\text{alt}}^{(n)}$ by

$$\alpha(v_1, \ldots, v_n) = \tau(S^{-1}v_1, \ldots, S^{-1}v_n)$$

for $v_1, \ldots, v_n \in V$. Suppose $w_1, \ldots, w_n \in W$. Then

$$\begin{aligned}\tau_{S^{-1}TS}(w_1, \ldots, w_n) &= \tau(S^{-1}TSw_1, \ldots, S^{-1}TSw_n)\\ &= \alpha(TSw_1, \ldots, TSw_n)\\ &= \alpha_T(Sw_1, \ldots, Sw_n)\\ &= (\det T)\alpha(Sw_1, \ldots, Sw_n)\\ &= (\det T)\tau(w_1, \ldots, w_n).\end{aligned}$$

The equation above and the definition of the determinant of the operator $S^{-1}TS$ imply that $\det(S^{-1}TS) = \det T$. □

For the special case in which $V = \mathbb{F}^n$ and $e_1, \dots, e_n$ is the standard basis of $\mathbb{F}^n$, the next result is true by the definition of the determinant of a matrix. The left side of the equation in the next result does not depend on a choice of basis, which means that the right side is independent of the choice of basis.

Determinant of operator equals determinant of its matrix Result 9.53

Suppose $T \in \text{End}(V)$ and $e_1, \dots, e_n$ is a basis of V. Then

$$\det T = \det \text{Mat}\big(T, (e_1, \dots, e_n)\big).$$

Proof. Let $f_1, \dots, f_n$ be the standard basis of $\mathbb{F}^n$. Let $S: \mathbb{F}^n \to V$ be the linear map such that $Sf_k = e_k$ for each $k = 1, \dots, n$. Thus $\text{Mat}\big(S, (f_1, \dots, f_n), (e_1, \dots, e_n)\big)$ and $\text{Mat}\big(S^{-1}, (e_1, \dots, e_n), (f_1, \dots, f_n)\big)$ both equal the n-by-n identity matrix. Hence

$$\text{Mat}\big(S^{-1}TS, (f_1, \dots, f_n)\big) = \text{Mat}\big(T, (e_1, \dots, e_n)\big), \tag{9.54}$$

as follows from two applications of 3.43. Thus

$$\begin{aligned} \det T &= \det(S^{-1}TS) \\ &= \det \text{Mat}\big(S^{-1}TS, (f_1, \dots, f_n)\big) \\ &= \det \text{Mat}\big(T, (e_1, \dots, e_n)\big), \end{aligned}$$

where the first line comes from 9.52, the second line comes from the definition of the determinant of a matrix, and the third line follows from 9.54. □

The next result gives a more intuitive way to think about determinants than the definition or the formula in 9.46. We could make the characterization in the result below the definition of the determinant of an operator on a finite-dimensional complex vector space, with the current definition then becoming a consequence of that definition.

If $\mathbb{F} = \mathbb{C}$, then determinant equals product of eigenvalues Result 9.55

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Then $\det T$ equals the product of the eigenvalues of T, with each eigenvalue included as many times as its multiplicity.

Proof. There is a basis of V with respect to which T has an upper-triangular matrix with the diagonal entries of the matrix consisting of the eigenvalues of T, with each eigenvalue included as many times as its multiplicity—see 8.37. Thus 9.53 and 9.48 imply that $\det T$ equals the product of the eigenvalues of T, with each eigenvalue included as many times as its multiplicity. □

As the next result shows, the determinant interacts nicely with the transpose of a square matrix, with the dual of an operator, and with the adjoint of an operator on an inner product space.

Determinant of transpose, dual, or adjoint — Result 9.56

(a) Suppose A is a square matrix. Then $\det A^t = \det A$.

(b) Suppose $T \in \text{End}(V)$. Then $\det T' = \det T$.

(c) Suppose V is an inner product space and $T \in \text{End}(V)$. Then

$$\det(T^*) = \overline{\det T}.$$

Proof.

(a) Let n be a positive integer. Define $\alpha\colon (\mathbb{F}^n)^n \to \mathbb{F}$ by

$$\alpha(v_1, \dots, v_n) = \det\Big(\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^t\Big)$$

for all $v_1, \dots, v_n \in \mathbb{F}^n$. The formula in 9.46 for the determinant of a matrix shows that α is an n-linear form on $\mathbb{F}^n$.

Suppose $v_1, \dots, v_n \in \mathbb{F}^n$ and $v_j = v_k$ for some $j \neq k$. If B is an n-by-n matrix, then $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^t B$ cannot equal the identity matrix because row j and row k of $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^t B$ are equal. Thus $\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^t$ is not invertible, which implies that $\alpha(v_1, \dots, v_n) = 0$. Hence α is an alternating n-linear form on $\mathbb{F}^n$.

Note that α applied to the standard basis of $\mathbb{F}^n$ equals 1. Because the vector space of alternating n-linear forms on $\mathbb{F}^n$ has dimension one (by 9.37), this implies that α is the determinant function. Thus (a) holds.

(b) The equation $\det T' = \det T$ follows from (a) and 9.53 and 3.132.

(c) Pick an orthonormal basis of V. The matrix of T^* with respect to that basis is the conjugate transpose of the matrix of T with respect to that basis (by 7.9). Thus 9.53, 9.46, and (a) imply that $\det(T^*) = \overline{\det T}$. □

Helpful results in evaluating determinants — Result 9.57

(a) If either two columns or two rows of a square matrix are equal, then the determinant of the matrix equals 0.

(b) Suppose A is a square matrix and B is the matrix obtained from A by swapping either two columns or two rows. Then $\det A = -\det B$.

(c) If one column or one row of a square matrix is multiplied by a scalar, then the value of the determinant is multiplied by the same scalar.

(d) If a scalar multiple of one column of a square matrix is added to another column, then the value of the determinant is unchanged.

(e) If a scalar multiple of one row of a square matrix is added to another row, then the value of the determinant is unchanged.

Proof. All the assertions in this result follow from the result that the maps $v_1, \dots, v_n \mapsto \det\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ and $v_1, \dots, v_n \mapsto \det\begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}^t$ are both alternating n-linear forms on $\mathbb{F}^n$ [see 9.45 and 9.56(a)].

For example, to prove (d) suppose $v_1, \dots, v_n \in \mathbb{F}^n$ and $c \in \mathbb{F}$. Then

$$\begin{aligned}
&\det\begin{pmatrix} v_1 + cv_2 & v_2 & \cdots & v_n \end{pmatrix} \\
&\quad = \det\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} + c \det\begin{pmatrix} v_2 & v_2 & v_3 & \cdots & v_n \end{pmatrix} \\
&\quad = \det\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix},
\end{aligned}$$

where the first equation follows from the multilinearity property and the second equation follows from the alternating property. The equation above shows that adding a multiple of the second column to the first column does not change the value of the determinant. The same conclusion holds for any two columns. Thus (d) holds.

The proof of (e) follows from (d) and from 9.56(a). The proofs of (a), (b), and (c) use similar tools and are left to the reader. □

For matrices whose entries are concrete numbers, the result above leads to a much faster way to evaluate the determinant than direct application of the formula in 9.46. Specifically, apply the Gaussian elimination procedure of swapping rows [by 9.57(b), this changes the determinant by a factor of -1], multiplying a row by a nonzero constant [by 9.57(c), this changes the determinant by the same constant], and adding a multiple of one row to another row [by 9.57(e), this does not change the determinant] to produce an upper-triangular matrix, whose determinant is the product of the diagonal entries (by 9.48). If your software keeps track of the number of row swaps and of the constants used when multiplying a row by a constant, then the determinant of the original matrix can be computed.

Because a number $\lambda \in \mathbb{F}$ is an eigenvalue of an operator $T \in \operatorname{End}(V)$ if and only if $\det(\lambda I - T) = 0$ (by 9.51), you may be tempted to think that one way to find eigenvalues quickly is to choose a basis of V, let $A = \operatorname{Mat}(T)$, evaluate $\det(\lambda I - A)$, and then solve the equation $\det(\lambda I - A) = 0$ for λ. However, that procedure is rarely efficient, except when $\dim V = 2$ (or when $\dim V$ equals 3 or 4 if you are willing to use the cubic or quartic formulas). One problem is that the procedure described in the paragraph above for evaluating a determinant does not work when the matrix includes a symbol (such as the λ in $\lambda I - A$). This problem arises because decisions need to be made in the Gaussian elimination procedure about whether certain quantities equal 0, and those decisions become complicated in expressions involving a symbol λ.

Recall that an operator on a finite-dimensional inner product space is unitary if it preserves norms (see 7.51 and the paragraph following it). Every eigenvalue of a unitary operator has absolute value 1 (by 7.54). Thus the product of the eigenvalues of a unitary operator has absolute value 1. Hence (at least in the case $\mathbb{F} = \mathbb{C}$) the determinant of a unitary operator has absolute value 1 (by 9.55). The next result gives a proof that works without the assumption that $\mathbb{F} = \mathbb{C}$.

Result 9.58

Every unitary operator has determinant with absolute value 1

Suppose V is an inner product space and $S \in \operatorname{End}(V)$ is a unitary operator. Then $|\det S| = 1$.

Proof. Because S is unitary, $I = S^*S$ (see 7.53). Thus

$$1 = \det(S^*S) = (\det S^*)(\det S) = (\overline{\det S})(\det S) = |\det S|^2,$$

where the second equality comes from 9.49(a) and the third equality comes from 9.56(c). The equation above implies that $|\det S| = 1$. □

The determinant of a positive operator on an inner product space meshes well with the analogy that such operators correspond to the nonnegative real numbers.

Every positive operator has nonnegative determinant **Result 9.59**

Suppose V is an inner product space and $T \in \text{End}(V)$ is a positive operator. Then $\det T \geq 0$.

Proof. By the spectral theorem (7.29 or 7.31), V has an orthonormal basis consisting of eigenvectors of T. Thus by the last bullet point of 9.42, $\det T$ equals a product of the eigenvalues of T, possibly with repetitions. Each eigenvalue of T is a nonnegative number (by 7.38). Thus we conclude that $\det T \geq 0$. □

Suppose V is an inner product space and $T \in \text{End}(V)$. Recall that the list of nonnegative square roots of the eigenvalues of T^*T (each included as many times as its multiplicity) is called the list of singular values of T (see Section 7E).

$|\det T|$ = product of singular values of T **Result 9.60**

Suppose V is an inner product space and $T \in \text{End}(V)$. Then

$$|\det T| = \sqrt{\det(T^*T)} = \text{product of singular values of } T.$$

Proof. We have

$$|\det T|^2 = (\overline{\det T})(\det T) = \big(\det(T^*)\big)(\det T) = \det(T^*T),$$

where the middle equality comes from 9.56(c) and the last equality comes from 9.49(a). Taking square roots of both sides of the equation above shows that $|\det T| = \sqrt{\det(T^*T)}$.

Let $s_1, \ldots, s_n$ denote the list of singular values of T. Thus $s_1^2, \ldots, s_n^2$ is the list of eigenvalues of T^*T (with appropriate repetitions), corresponding to an orthonormal basis of V consisting of eigenvectors of T^*T. Hence the last bullet point of 9.42 implies that

$$\det(T^*T) = s_1^2 \cdots s_n^2.$$

Thus $|\det T| = s_1 \cdots s_n$, as desired. □

An operator T on a real inner product space changes volume by a factor of the product of the singular values (by 7.111). Thus the next result follows immediately from 7.111 and 9.60. This result explains why the absolute value of a determinant appears in the change of variables formula in multivariable calculus.

T changes volume by factor of $|\det T|$ — Result 9.61

Suppose $T \in \text{End}(\mathbb{R}^n)$ and $\Omega \subseteq \mathbb{R}^n$. Then

$$\text{volume } T(\Omega) = |\det T|(\text{volume } \Omega).$$

For operators on finite-dimensional complex vector spaces, we now connect the determinant to a polynomial that we have previously seen.

If $\mathbb{F} = \mathbb{C}$, then characteristic polynomial of T equals $\det(zI - T)$ — Result 9.62

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T, and let $d_1, \dots, d_m$ denote their multiplicities. Then

$$\det(zI - T) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}.$$

Proof. There exists a basis of V with respect to which T has an upper-triangular matrix with each λ_k appearing on the diagonal exactly d_k times (by 8.37). With respect to this basis, $zI - T$ has an upper-triangular matrix with $z - \lambda_k$ appearing on the diagonal exactly d_k times for each k. Thus 9.48 gives the desired equation. □

Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \text{End}(V)$. The characteristic polynomial of T was defined in 8.26 as the polynomial on the right side of the equation in 9.62. We did not previously define the characteristic polynomial of an operator on a finite-dimensional real vector space because such operators may have no eigenvalues, making a definition using the right side of the equation in 9.62 inappropriate.

We now present a new definition of the characteristic polynomial, motivated by 9.62. This new definition is valid for both real and complex vector spaces. The equation in 9.62 shows that this new definition is equivalent to our previous definition when $\mathbb{F} = \mathbb{C}$ (8.26).

Characteristic polynomial — Definition 9.63

Suppose $T \in \text{End}(V)$. The polynomial defined by

$$z \mapsto \det(zI - T)$$

is called the *characteristic polynomial* of T.

The formula in 9.46 shows that the characteristic polynomial of an operator $T \in \text{End}(V)$ is a monic polynomial of degree $\dim V$. The zeros in $\mathbb{F}$ of the characteristic polynomial of T are exactly the eigenvalues of T (by 9.51).

Previously we proved the Cayley–Hamilton theorem (8.29) in the complex case. Now we can extend that result to operators on real vector spaces.

Cayley–Hamilton theorem — Result 9.64

Suppose $T \in \text{End}(V)$ and q is the characteristic polynomial of T. Then $q(T) = 0$.

Proof. If $\mathbb{F} = \mathbb{C}$, then the equation $q(T) = 0$ follows from 9.62 and 8.29.

Now suppose $\mathbb{F} = \mathbb{R}$. Fix a basis of V, and let A be the matrix of T with respect to this basis. Let S be the operator on $\mathbb{C}^{\dim V}$ such that the matrix of S (with respect to the standard basis of $\mathbb{C}^{\dim V}$) is A. For all $z \in \mathbb{R}$ we have

$$q(z) = \det(zI - T) = \det(zI - A) = \det(zI - S).$$

Thus q is the characteristic polynomial of S. The case $\mathbb{F} = \mathbb{C}$ (first sentence of this proof) now implies that $0 = q(S) = q(A) = q(T)$. □

The Cayley–Hamilton theorem (9.64) implies that the characteristic polynomial of an operator $T \in \text{End}(V)$ is a polynomial multiple of the minimal polynomial of T (by 5.29). Thus if the degree of the minimal polynomial of T equals $\dim V$, then the characteristic polynomial of T equals the minimal polynomial of T. This happens for a very large percentage of operators, including over 99.999% of 4-by-4 matrices with integer entries in $[-100, 100]$ (see the paragraph following 5.25).

The last sentence in our next result was previously proved in the complex case (see 8.54). Now we can give a proof that works on both real and complex vector spaces.

Characteristic polynomial, trace, and determinant — **Result 9.65**

Suppose $T \in \text{End}(V)$. Let $n = \dim V$. Then the characteristic polynomial of T can be written as

$$z^n - (\text{tr}\, T)z^{n-1} + \cdots + (-1)^n(\det T).$$

Proof. The constant term of a polynomial function of z is the value of the polynomial when $z = 0$. Thus the constant term of the characteristic polynomial of T equals $\det(-T)$, which equals $(-1)^n \det T$ (by the third bullet point of 9.42).

Fix a basis of V, and let A be the matrix of T with respect to this basis. The matrix of $zI - T$ with respect to this basis is $zI - A$. The term coming from the identity permutation $\{1, \ldots, n\}$ in the formula 9.46 for $\det(zI - A)$ is

$$(z - A_{1,1}) \cdots (z - A_{n,n}).$$

The coefficient of z^{n-1} in the expression above is $-(A_{1,1} + \cdots + A_{n,n})$, which equals $-\text{tr}\, T$. The terms in the formula for $\det(zI - A)$ coming from other elements of $\text{perm}\, n$ contain at most $n - 2$ factors of the form $z - A_{k,k}$ and thus do not contribute to the coefficient of z^{n-1} in the characteristic polynomial of T. □

The next result was proved by Jacques Hadamard (1865–1963) in 1893.

In the result below, think of the columns of the n-by-n matrix A as elements of $\mathbb{F}^n$. The norms appearing below then arise from the standard inner product on $\mathbb{F}^n$. Recall that the notation $R_{\cdot,k}$ in the proof below means the kth column of the matrix R (as was defined in 3.44).

Hadamard's inequality — **Result 9.66**

Result 9.66 *continued*

Suppose A is an n-by-n matrix. Let $v_1, \dots, v_n$ denote the columns of A. Then

$$|\det A| \le \prod_{k=1}^{n} \|v_k\|.$$

Proof. If A is not invertible, then $\det A = 0$ and hence the desired inequality holds in this case.

Thus assume that A is invertible. The QR factorization (7.58) tells us that there exist a unitary matrix Q and an upper-triangular matrix R whose diagonal contains only positive numbers such that $A = QR$. We have

$$\begin{aligned}
|\det A| &= |\det Q| \, |\det R| \\
&= |\det R| \\
&= \prod_{k=1}^{n} R_{k,k} \\
&\le \prod_{k=1}^{n} \|R_{\cdot,k}\| \\
&= \prod_{k=1}^{n} \|QR_{\cdot,k}\| \\
&= \prod_{k=1}^{n} \|v_k\|,
\end{aligned}$$

where the first line comes from 9.49(b), the second line comes from 9.58, the third line comes from 9.48, and the fifth line holds because Q is an isometry. □

To give a geometric interpretation to Hadamard's inequality, suppose $\mathbb{F} = \mathbb{R}$. Let $T \in \text{End}(\mathbb{R}^n)$ be the operator such that $Te_k = v_k$ for each $k = 1, \dots, n$, where $e_1, \dots, e_n$ is the standard basis of $\mathbb{R}^n$. Then T maps the box $P(e_1, \dots, e_n)$ onto the parallelepiped $P(v_1, \dots, v_n)$ [see 7.102 and 7.105 for a review of this notation and terminology]. Because the box $P(e_1, \dots, e_n)$ has volume 1, this implies (by 9.61) that the parallelepiped $P(v_1, \dots, v_n)$ has volume $|\det T|$, which equals $|\det A|$. Thus Hadamard's inequality above can be interpreted to say that among all parallelepipeds whose edges have lengths $\|v_1\|, \dots, \|v_n\|$, the ones with largest volume have orthogonal edges (and thus have volume $\prod_{k=1}^{n} \|v_k\|$).

For a necessary and sufficient condition for Hadamard's inequality to be an equality, see Exercise 18.

The matrix in the next result is called the *Vandermonde matrix*. Vandermonde matrices have important applications in polynomial interpolation, the discrete Fourier transform, and other areas of mathematics. The proof of the next result is a nice illustration of the power of switching between matrices and linear maps.

Result 9.67

Determinant of Vandermonde matrix

Result 9.67 *continued*

Suppose $n > 1$ and $\beta_1, \dots, \beta_n \in \mathbb{F}$. Then

$$\det \begin{pmatrix} 1 & \beta_1 & \beta_1^2 & \cdots & \beta_1^{n-1} \\ 1 & \beta_2 & \beta_2^2 & \cdots & \beta_2^{n-1} \\ & & \ddots & & \\ 1 & \beta_n & \beta_n^2 & \cdots & \beta_n^{n-1} \end{pmatrix} = \prod_{1 \le j < k \le n} (\beta_k - \beta_j).$$

Proof. Let $1, z, \dots, z^{n-1}$ be the standard basis of $\mathbb{F}[X]_{\le n-1}$ and let $e_1, \dots, e_n$ denote the standard basis of $\mathbb{F}^n$. Define a linear map $S : \mathbb{F}[X]_{\le n-1} \to \mathbb{F}^n$ by

$$Sp = \big(p(\beta_1), \dots, p(\beta_n)\big).$$

Let A denote the Vandermonde matrix shown in the statement of this result. Note that

$$A = \operatorname{Mat}\big(S, (1, z, \dots, z^{n-1}), (e_1, \dots, e_n)\big).$$

Let $T : \mathbb{F}[X]_{\le n-1} \to \mathbb{F}[X]_{\le n-1}$ be the operator on $\mathbb{F}[X]_{\le n-1}$ such that $T1 = 1$ and

$$Tz^k = (z - \beta_1)(z - \beta_2) \cdots (z - \beta_k)$$

for $k = 1, \dots, n-1$. Let $B = \operatorname{Mat}\big(T, (1, z, \dots, z^{n-1}), (1, z, \dots, z^{n-1})\big)$. Then B is an upper-triangular matrix all of whose diagonal entries equal 1. Thus $\det B = 1$ (by 9.48).

Let $C = \operatorname{Mat}\big(ST, (1, z, \dots, z^{n-1}), (e_1, \dots, e_n)\big)$. Thus $C = AB$ (by 3.81), which implies that

$$\det A = (\det A)(\det B) = \det C.$$

The definitions of C, S, and T show that C equals

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & \beta_2 - \beta_1 & 0 & \cdots & 0 \\ 1 & \beta_3 - \beta_1 & (\beta_3 - \beta_1)(\beta_3 - \beta_2) & \cdots & 0 \\ & & & \ddots & \\ 1 & \beta_n - \beta_1 & (\beta_n - \beta_1)(\beta_n - \beta_2) & \cdots & (\beta_n - \beta_1)(\beta_n - \beta_2) \cdots (\beta_n - \beta_{n-1}) \end{pmatrix}.$$

Now $\det A = \det C = \prod_{1 \le j < k \le n} (\beta_k - \beta_j)$, where we have used 9.56(a) and 9.48. □

Exercises

9C.1 Prove or give a counterexample: $S, T \in \operatorname{End}(V) \implies \det(S + T) = \det S + \det T$.

9C.2 Suppose the first column of a square matrix A consists of all zeros except possibly the first entry $A_{1,1}$. Let B be the matrix obtained from A by deleting the first row and the first column of A. Show that $\det A = A_{1,1} \det B$.

9C.3 Suppose $T \in \text{End}(V)$ is nilpotent. Prove that $\det(I + T) = 1$.

9C.4 Suppose $S \in \text{End}(V)$. Prove that S is unitary if and only if $|\det S| = \|S\| = 1$.

9C.5 Suppose A is a block upper-triangular matrix

$$A = \begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_k along the diagonal is a square matrix. Prove that

$$\det A = (\det A_1) \cdots (\det A_m).$$

9C.6 Suppose $A = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$ is an n-by-n matrix, with v_k denoting the kth column of A. Show that if $(m_1, \ldots, m_n) \in \text{perm}\, n$, then

$$\det \begin{pmatrix} v_{m_1} & \cdots & v_{m_n} \end{pmatrix} = \big(\text{sign}(m_1, \ldots, m_n)\big) \det A.$$

9C.7 Suppose $T \in \text{End}(V)$ 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 $z \in \mathbb{F}$.

9C.8 Suppose $T \in \text{End}(V)$ is an operator with no eigenvalues (which implies that $\mathbb{F} = \mathbb{R}$). Prove that $\det T > 0$.

9C.9 Suppose that V is a real vector space of even dimension, $T \in \text{End}(V)$, and $\det T < 0$. Prove that T has at least two distinct eigenvalues.

9C.10 Suppose V is a real vector space of odd dimension and $T \in \text{End}(V)$. Without using the minimal polynomial, prove that T has an eigenvalue.

This result was previously proved without using determinants or the characteristic polynomial—see 5.34.

9C.11 Prove or give a counterexample: If $\mathbb{F} = \mathbb{R}$, $T \in \text{End}(V)$, and $\det T > 0$, then T has a square root.

If $\mathbb{F} = \mathbb{C}$, $T \in \text{End}(V)$, and $\det T \neq 0$, then T has a square root (see 8.41).

9C.12 Suppose $S, T \in \text{End}(V)$ and S is invertible. Define $p: \mathbb{F} \to \mathbb{F}$ by

$$p(z) = \det(zS - T).$$

Prove that p is a polynomial of degree $\dim V$ and that the coefficient of $z^{\dim V}$ in this polynomial is $\det S$.

9C.13 Suppose $\mathbb{F} = \mathbb{C}$, $T \in \text{End}(V)$, and $n = \dim V > 2$. Let $\lambda_1, \ldots, \lambda_n$ denote the eigenvalues of T, with each eigenvalue included as many times as its multiplicity.

(a) Find a formula for the coefficient of z^{n-2} in the characteristic polynomial of T in terms of $\lambda_1, \ldots, \lambda_n$.

(b) Find a formula for the coefficient of z in the characteristic polynomial of T in terms of $\lambda_1, \ldots, \lambda_n$.

9C.14 Suppose V is an inner product space and T is a positive operator on V. Prove that

$$\det \sqrt{T} = \sqrt{\det T}.$$

9C.15 Suppose V is an inner product space and $T \in \text{End}(V)$. Use the polar decomposition to give a proof that

$$|\det T| = \sqrt{\det(T^*T)}$$

that is different from the proof given earlier (see 9.60).

9C.16 Suppose $T \in \text{End}(V)$. Define $g: \mathbb{F} \to \mathbb{F}$ by $g(x) = \det(I + xT)$. Show that $g'(0) = \text{tr}\, T$.

Look for a clean solution to this exercise, without using the explicit but complicated formula for the determinant of a matrix.

9C.17 Suppose a, b, c are positive numbers. Find the volume of the ellipsoid

$$\left\{(x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} < 1\right\}$$

by finding a set $\Omega \subseteq \mathbb{R}^3$ whose volume you know and an operator T on $\mathbb{R}^3$ such that $T(\Omega)$ equals the ellipsoid above.

9C.18 Suppose that A is an invertible square matrix. Prove that Hadamard's inequality (9.66) is an equality if and only if each column of A is orthogonal to the other columns.

9C.19 Suppose V is an inner product space, $e_1, \dots, e_n$ is an orthonormal basis of V, and $T \in \text{End}(V)$ is a positive operator.

(a) Prove that $\det T \le \prod_{k=1}^{n} \langle Te_k, e_k \rangle$.

(b) Prove that if T is invertible, then the inequality in (a) is an equality if and only if e_k is an eigenvector of T for each $k = 1, \dots, n$.

9C.20 Suppose A is an n-by-n matrix, and suppose c is such that $|A_{j,k}| \le c$ for all $j, k \in \{1, \dots, n\}$. Prove that

$$|\det A| \le c^n n^{n/2}.$$

The formula for the determinant of a matrix (9.46) shows that $|\det A| \le c^n n!$. However, the estimate given by this exercise is much better. For example, if $c = 1$ and $n = 100$, then $c^n n! \approx 10^{158}$, but the estimate given by this exercise is the much smaller number 10^{100}. If n is an integer power of 2, then the inequality above is sharp and cannot be improved.

9C.21 Suppose n is a positive integer and $\delta : \mathbb{C}^{n,n} \to \mathbb{C}$ is a function such that

$$\delta(AB) = \delta(A) \cdot \delta(B)$$

for all $A, B \in \mathbb{C}^{n,n}$ and $\delta(A)$ equals the product of the diagonal entries of A for each diagonal matrix $A \in \mathbb{C}^{n,n}$. Prove that

$$\delta(A) = \det A$$

for all $A \in \mathbb{C}^{n,n}$.

Recall that $\mathbb{C}^{n,n}$ denotes the set of n-by-n matrices with entries in $\mathbb{C}$. This exercise shows that the determinant is the unique function defined on square matrices that is multiplicative and has the desired behavior on diagonal matrices. This result is analogous to Exercise 10 in Section 8D, which shows that the trace is uniquely determined by its algebraic properties.

I find that in my own elementary lectures, I have, for pedagogical reasons, pushed determinants more and more into the background. Too often I have had the experience that, while the students acquired facility with the formulas, which are so useful in abbreviating long expressions, they often failed to gain familiarity with their meaning, *and skill in manipulation prevented the student from going into all the details of the subject and so gaining a mastery.*

—*Elementary Mathematics from an Advanced Standpoint: Geometry*, Felix Klein

9D Tensor Products

Tensor Product of Two Vector Spaces

The motivation for our next topic comes from wanting to form the product of a vector $v \in V$ and a vector $w \in W$. This product will be denoted by $v \otimes w$, pronounced

"v tensor w", and will be an element of some new vector space called $V \otimes W$ (also pronounced "V tensor W").

We already have a vector space $V \times W$ (see Section 3E), called the product of V and W. However, $V \times W$ will not serve our purposes here because it does not provide a natural way to multiply an element of V by an element of W. We would like our tensor product to satisfy some of the usual properties of multiplication. For example, we would like the distributive property to be satisfied, meaning that if $v_1, v_2, v \in V$ and $w_1, w_2, w \in W$, then

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w \quad \text{and} \quad v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2.$$

We would also like scalar multiplication to interact well with this new multiplication, meaning that

$$\lambda(v \otimes w) = (\lambda v) \otimes w = v \otimes (\lambda w)$$

for all $\lambda \in \mathbb{F}$, $v \in V$, and $w \in W$.

To produce ⊗ in TEX, type \otimes.

Furthermore, it would be nice if each basis of V when combined with each basis of W produced a basis of $V \otimes W$. Specifically, if $e_1, \dots, e_m$ is a basis of V and $f_1, \dots, f_n$ is a basis of W, then we would like a list (in any order) consisting of $e_j \otimes f_k$, as j ranges from 1 to m and k ranges from 1 to n, to be a basis of $V \otimes W$. This implies that $\dim(V \otimes W)$ should equal $(\dim V)(\dim W)$. Recall that $\dim(V \times W) = \dim V + \dim W$ (see 3.92), which shows that the product $V \times W$ will not serve our purposes here.

To produce a vector space whose dimension is $(\dim V)(\dim W)$ in a natural fashion from V and W, we look at the vector space of bilinear functionals, as defined below.

Definition 9.68 **Bilinear functional on $V \times W$, the vector space $\mathcal{B}(V, W)$**

- A *bilinear functional* on $V \times W$ is a function $\beta: V \times W \to \mathbb{F}$ such that $v \mapsto \beta(v, w)$ is a linear functional on V for each $w \in W$ and $w \mapsto \beta(v, w)$ is a linear functional on W for each $v \in V$.
- The vector space of bilinear functionals on $V \times W$ is denoted by $\mathcal{B}(V, W)$.

If $W = V$, then a bilinear functional on $V \times W$ is a bilinear form; see 9.1.

The operations of addition and scalar multiplication on $\mathcal{B}(V, W)$ are defined to be the usual operations of addition and scalar multiplication of functions. As you can verify, these operations make $\mathcal{B}(V, W)$ into a vector space whose additive identity is the zero function from $V \times W$ to $\mathbb{F}$.

Example 9.69 **Bilinear functionals**

- Suppose $\varphi \in V'$ and $\tau \in W'$. Define $\beta: V \times W \to \mathbb{F}$ by $\beta(v, w) = \varphi(v)\tau(w)$. Then β is a bilinear functional on $V \times W$.
- Suppose $v \in V$ and $w \in W$. Define $\beta: V' \times W' \to \mathbb{F}$ by $\beta(\varphi, \tau) = \varphi(v)\tau(w)$. Then β is a bilinear functional on $V' \times W'$.

Example 9.69
continued

- Define $\beta: V \times V' \to \mathbb{F}$ by $\beta(v, \varphi) = \varphi(v)$. Then β is a bilinear functional on $V \times V'$.
- Suppose $\varphi \in V'$. Define $\beta: V \times \text{End}(V) \to \mathbb{F}$ by $\beta(v, T) = \varphi(Tv)$. Then β is a bilinear functional on $V \times \text{End}(V)$.
- Suppose m and n are positive integers. Define $\beta: \mathbb{F}^{m,n} \times \mathbb{F}^{n,m} \to \mathbb{F}$ by $\beta(A, B) = \text{tr}(AB)$. Then β is a bilinear functional on $\mathbb{F}^{m,n} \times \mathbb{F}^{n,m}$.

Dimension of the vector space of bilinear functionals — Result 9.70

$$\dim \mathcal{B}(V, W) = (\dim V)(\dim W).$$

Proof. Let $e_1, \ldots, e_m$ be a basis of V and $f_1, \ldots, f_n$ be a basis of W. For a bilinear functional $\beta \in \mathcal{B}(V, W)$, let $\text{Mat}(\beta)$ be the m-by-n matrix whose entry in row j, column k is $\beta(e_j, f_k)$. The map $\beta \mapsto \text{Mat}(\beta)$ is a linear map of $\mathcal{B}(V, W)$ into $\mathbb{F}^{m,n}$.

For a matrix $C \in \mathbb{F}^{m,n}$, define a bilinear functional β_C on $V \times W$ by

$$\beta_C(a_1 e_1 + \cdots + a_m e_m, b_1 f_1 + \cdots + b_n f_n) = \sum_{k=1}^{n} \sum_{j=1}^{m} C_{j,k} a_j b_k$$

for $a_1, \ldots, a_m, b_1, \ldots, b_n \in \mathbb{F}$.

The linear map $\beta \mapsto \text{Mat}(\beta)$ from $\mathcal{B}(V, W)$ to $\mathbb{F}^{m,n}$ and the linear map $C \mapsto \beta_C$ from $\mathbb{F}^{m,n}$ to $\mathcal{B}(V, W)$ are inverses of each other because $\beta_{\text{Mat}(\beta)} = \beta$ for all $\beta \in \mathcal{B}(V, W)$ and $\text{Mat}(\beta_C) = C$ for all $C \in \mathbb{F}^{m,n}$, as you should verify.

Thus both maps are isomorphisms and the two spaces that they connect have the same dimension. Hence $\dim \mathcal{B}(V, W) = \dim \mathbb{F}^{m,n} = mn = (\dim V)(\dim W)$. □

Several different definitions of $V \otimes W$ appear in the mathematical literature. These definitions are equivalent to each other, at least in the finite-dimensional context, because any two vector spaces of the same dimension are isomorphic.

The result above states that $\mathcal{B}(V, W)$ has the dimension that we seek, as do $\text{Hom}(V, W)$ and $\mathbb{F}^{\dim V, \dim W}$. Thus it may be tempting to define $V \otimes W$ to be $\mathcal{B}(V, W)$ or $\text{Hom}(V, W)$ or $\mathbb{F}^{\dim V, \dim W}$. However, none of those definitions would lead to a basis-free definition of $v \otimes w$ for $v \in V$ and $w \in W$.

The following definition, while it may seem a bit strange and abstract at first, has the huge advantage that it defines $v \otimes w$ in a basis-free fashion. We define $V \otimes W$ to be the vector space of bilinear functionals on $V' \times W'$ instead of the more tempting choice of the vector space of bilinear functionals on $V \times W$.

Tensor product, $V \otimes W$, $v \otimes w$ — Definition 9.71

- The *tensor product* $V \otimes W$ is defined to be $\mathcal{B}(V', W')$.
- For $v \in V$ and $w \in W$, the *tensor product* $v \otimes w$ is the element of $V \otimes W$

Definition 9.71 *continued*

defined by

$$(v \otimes w)(\varphi, \tau) = \varphi(v)\tau(w)$$

for all $(\varphi, \tau) \in V' \times W'$.

We can quickly prove that the definition of $V \otimes W$ gives it the desired dimension.

Result 9.72 **Dimension of the tensor product of two vector spaces**

$$\dim(V \otimes W) = (\dim V)(\dim W).$$

Proof. Because a vector space and its dual have the same dimension (by 3.111), we have $\dim V' = \dim V$ and $\dim W' = \dim W$. Thus 9.70 tells us that the dimension of $\mathcal{B}(V', W')$ equals $(\dim V)(\dim W)$. □

To understand the definition of the tensor product $v \otimes w$ of two vectors $v \in V$ and $w \in W$, focus on the kind of object it is. An element of $V \otimes W$ is a bilinear functional on $V' \times W'$, and in particular it is a function from $V' \times W'$ to $\mathbb{F}$. Thus for each element of $V' \times W'$, it should produce an element of $\mathbb{F}$. The definition above has this behavior, because $v \otimes w$ applied to a typical element (φ, τ) of $V' \times W'$ produces the number $\varphi(v)\tau(w)$.

The somewhat abstract nature of $v \otimes w$ should not matter. The important point is the behavior of these objects. The next result shows that tensor products of vectors have the desired bilinearity properties.

Result 9.73 **Bilinearity of tensor product**

Suppose $v, v_1, v_2 \in V$ and $w, w_1, w_2 \in W$ and $\lambda \in \mathbb{F}$. Then

$$(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w \quad \text{and} \quad v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$$

and

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

Proof. Suppose $(\varphi, \tau) \in V' \times W'$. Then

$$\begin{aligned}\big((v_1 + v_2) \otimes w\big)(\varphi, \tau) &= \varphi(v_1 + v_2)\tau(w) \\ &= \varphi(v_1)\tau(w) + \varphi(v_2)\tau(w) \\ &= (v_1 \otimes w)(\varphi, \tau) + (v_2 \otimes w)(\varphi, \tau) \\ &= (v_1 \otimes w + v_2 \otimes w)(\varphi, \tau).\end{aligned}$$

Thus $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$.

The other two equalities are proved similarly. □

Lists are, by definition, ordered. The order matters when, for example, we form the matrix of an operator with respect to a basis. For lists in this section with two

indices, such as $\{e_j \otimes f_k\}_{j=1,\dots,m;\,k=1,\dots,n}$ in the next result, the ordering does not matter and we do not specify it—just choose any convenient ordering.

The linear independence of elements of $V \otimes W$ in (a) of the result below captures the idea that there are no relationships among vectors in $V \otimes W$ other than the relationships that come from bilinearity of the tensor product (see 9.73) and the relationships that may be present due to linear dependence of a list of vectors in V or a list of vectors in W.

Result 9.74

Basis of $V \otimes W$

Suppose $e_1, \dots, e_m$ is a list of vectors in V and $f_1, \dots, f_n$ is a list of vectors in W.

(a) If $e_1, \dots, e_m$ and $f_1, \dots, f_n$ are both linearly independent lists, then

$$\{e_j \otimes f_k\}_{j=1,\dots,m;\,k=1,\dots,n}$$

is a linearly independent list in $V \otimes W$.

(b) If $e_1, \dots, e_m$ is a basis of V and $f_1, \dots, f_n$ is a basis of W, then the list $\{e_j \otimes f_k\}_{j=1,\dots,m;\,k=1,\dots,n}$ is a basis of $V \otimes W$.

Proof. To prove (a), suppose $e_1, \dots, e_m$ and $f_1, \dots, f_n$ are both linearly independent lists. This linear independence and the linear map lemma (3.4) imply that there exist $\varphi_1, \dots, \varphi_m \in V'$ and $\tau_1, \dots, \tau_n \in W'$ such that

$$\varphi_j(e_k) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases} \quad \text{and} \quad \tau_j(f_k) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k, \end{cases}$$

where $j, k \in \{1, \dots, m\}$ in the first equation and $j, k \in \{1, \dots, n\}$ in the second equation.

Suppose $\{a_{j,k}\}_{j=1,\dots,m;\,k=1,\dots,n}$ is a list of scalars such that

$$\sum_{k=1}^{n} \sum_{j=1}^{m} a_{j,k}(e_j \otimes f_k) = 0. \tag{9.75}$$

Note that $(e_j \otimes f_k)(\varphi_M, \tau_N)$ equals 1 if $j = M$ and $k = N$, and equals 0 otherwise. Thus applying both sides of 9.75 to (φ_M, τ_N) shows that $a_{M,N} = 0$, proving that $\{e_j \otimes f_k\}_{j=1,\dots,m;\,k=1,\dots,n}$ is linearly independent.

Now (b) follows from (a), the equation $\dim V \otimes W = (\dim V)(\dim W)$ [see 9.72], and the result that a linearly independent list of the right length is a basis (see 2.38). □

Every element of $V \otimes W$ is a finite sum of elements of the form $v \otimes w$, where $v \in V$ and $w \in W$, as implied by (b) in the result above. However, if $\dim V > 1$ and $\dim W > 1$, then Exercise 4 shows that

$$\{v \otimes w : (v, w) \in V \times W\} \neq V \otimes W.$$

Example 9.76 **Tensor product of element of $\mathbb{F}^m$ with element of $\mathbb{F}^n$**

Suppose m and n are positive integers. Let $e_1, \dots, e_m$ denote the standard basis of $\mathbb{F}^m$ and let $f_1, \dots, f_n$ denote the standard basis of $\mathbb{F}^n$. Suppose

$$v = (v_1, \dots, v_m) \in \mathbb{F}^m \quad \text{and} \quad w = (w_1, \dots, w_n) \in \mathbb{F}^n.$$

Then

$$\begin{aligned} v \otimes w &= \left(\sum_{j=1}^{m} v_j e_j\right) \otimes \left(\sum_{k=1}^{n} w_k f_k\right) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{m} (v_j w_k)(e_j \otimes f_k). \end{aligned}$$

Thus with respect to the basis $\{e_j \otimes f_k\}_{j=1,\dots,m;\, k=1,\dots,n}$ of $\mathbb{F}^m \otimes \mathbb{F}^n$ provided by 9.74(b), the coefficients of $v \otimes w$ are the numbers $\{v_j w_k\}_{j=1,\dots,m;\, k=1,\dots,n}$. If instead of writing these numbers in a list, we write them in an m-by-n matrix with $v_j w_k$ in row j, column k, then we can identify $v \otimes w$ with the m-by-n matrix

$$\begin{pmatrix} v_1 w_1 & \cdots & v_1 w_n \\ & \ddots & \\ v_m w_1 & \cdots & v_m w_n \end{pmatrix}.$$

See Exercises 5 and 6 for practice in using the identification from the example above.

We now define bilinear maps, which differ from bilinear functionals in that the target space can be an arbitrary vector space rather than just the scalar field.

Bilinear map **Definition 9.77**

A *bilinear map* from $V \times W$ to a vector space U is a function $\Gamma: V \times W \to U$ such that $v \mapsto \Gamma(v, w)$ is a linear map from V to U for each $w \in W$ and $w \mapsto \Gamma(v, w)$ is a linear map from W to U for each $v \in V$.

Example 9.78 **Bilinear maps**

- Every bilinear functional on $V \times W$ is a bilinear map from $V \times W$ to $\mathbb{F}$.
- The function $\Gamma: V \times W \to V \otimes W$ defined by $\Gamma(v, w) = v \otimes w$ is a bilinear map from $V \times W$ to $V \otimes W$ (by 9.73).
- The function $\Gamma: \text{End}(V) \times \text{End}(V) \to \text{End}(V)$ defined by $\Gamma(S, T) = ST$ is a bilinear map from $\text{End}(V) \times \text{End}(V)$ to $\text{End}(V)$.
- The function $\Gamma: V \times \text{Hom}(V, W) \to W$ defined by $\Gamma(v, T) = Tv$ is a bilinear map from $V \times \text{Hom}(V, W)$ to W.

Tensor products allow us to convert bilinear maps on $V \times W$ into linear maps on $V \otimes W$ (and vice versa), as shown by the next result. In the mathematical literature, (a) of the result below is called the "universal property" of tensor products.

Converting bilinear maps to linear maps Result 9.79

Suppose U is a vector space.

(a) Suppose $\Gamma: V \times W \to U$ is a bilinear map. Then there exists a unique linear map $\hat{\Gamma}: V \otimes W \to U$ such that

$$\hat{\Gamma}(v \otimes w) = \Gamma(v, w)$$

for all $(v, w) \in V \times W$.

(b) Conversely, suppose $T: V \otimes W \to U$ is a linear map. Then there exists a unique bilinear map $T^{\#}: V \times W \to U$ such that

$$T^{\#}(v, w) = T(v \otimes w)$$

for all $(v, w) \in V \times W$.

Proof. Let $e_1, \ldots, e_m$ be a basis of V and let $f_1, \ldots, f_n$ be a basis of W. By the linear map lemma (3.4) and 9.74(b), there exists a unique linear map $\hat{\Gamma}: V \otimes W \to U$ such that

$$\hat{\Gamma}(e_j \otimes f_k) = \Gamma(e_j, f_k)$$

for all $j \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, n\}$.

Now suppose $(v, w) \in V \times W$. There exist $a_1, \ldots, a_m, b_1, \ldots, b_n \in \mathbb{F}$ such that $v = a_1 e_1 + \cdots + a_m e_m$ and $w = b_1 f_1 + \cdots + b_n f_n$. Thus

$$\begin{aligned}
\hat{\Gamma}(v \otimes w) &= \hat{\Gamma}\Big(\sum_{k=1}^{n} \sum_{j=1}^{m} (a_j b_k)(e_j \otimes f_k)\Big) \\
&= \sum_{k=1}^{n} \sum_{j=1}^{m} a_j b_k \hat{\Gamma}(e_j \otimes f_k) \\
&= \sum_{k=1}^{n} \sum_{j=1}^{m} a_j b_k \Gamma(e_j, f_k) \\
&= \Gamma(v, w),
\end{aligned}$$

as desired, where the second line holds because $\hat{\Gamma}$ is linear, the third line holds by the definition of $\hat{\Gamma}$, and the fourth line holds because Γ is bilinear.

The uniqueness of the linear map $\hat{\Gamma}$ satisfying $\hat{\Gamma}(v \otimes w) = \Gamma(v, w)$ follows from 9.74(b), completing the proof of (a).

To prove (b), define a function $T^{\#}: V \times W \to U$ by $T^{\#}(v, w) = T(v \otimes w)$ for all $(v, w) \in V \times W$. The bilinearity of the tensor product (see 9.73) and the linearity of T imply that $T^{\#}$ is bilinear.

Clearly the choice of $T^{\#}$ that satisfies the conditions is unique. □

To prove 9.79(a), we could not just define $\hat{\Gamma}(v \otimes w) = \Gamma(v, w)$ for all $v \in V$ and $w \in W$ (and then extend $\hat{\Gamma}$ linearly to all of $V \otimes W$) because elements of $V \otimes W$ do not have unique representations as finite sums of elements of the form $v \otimes w$. Our proof used a basis of V and a basis of W to get around this problem.

Although our construction of $\hat{\Gamma}$ in the proof of 9.79(a) depended on a basis of V and a basis of W, the equation $\hat{\Gamma}(v \otimes w) = \Gamma(v, w)$ that holds for all $v \in V$ and $w \in W$ shows that $\hat{\Gamma}$ does not depend on the choice of bases for V and W.

Tensor Product of Inner Product Spaces

The result below features three inner products—one on $V \otimes W$, one on V, and one on W, although we use the same symbol $\langle \cdot, \cdot \rangle$ for all three inner products.

Result 9.80

Inner product on tensor product of two inner product spaces
Suppose V and W are inner product spaces. Then there is a unique inner product on $V \otimes W$ such that

$$\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle \langle w, x \rangle$$

for all $v, u \in V$ and $w, x \in W$.

Proof. Suppose $e_1, \dots, e_m$ is an orthonormal basis of V and $f_1, \dots, f_n$ is an orthonormal basis of W. Define an inner product on $V \otimes W$ by

$$\left\langle \sum_{k=1}^{n} \sum_{j=1}^{m} b_{j,k}\, e_j \otimes f_k,\ \sum_{k=1}^{n} \sum_{j=1}^{m} c_{j,k}\, e_j \otimes f_k \right\rangle = \sum_{k=1}^{n} \sum_{j=1}^{m} b_{j,k} \overline{c_{j,k}}. \tag{9.81}$$

The straightforward verification that 9.81 defines an inner product on $V \otimes W$ is left to the reader [use 9.74(b)].

Suppose that $v, u \in V$ and $w, x \in W$. Let $v_1, \dots, v_m \in \mathbb{F}$ be such that $v = v_1 e_1 + \cdots + v_m e_m$, with similar expressions for u, w, and x. Then

$$\begin{aligned}
\langle v \otimes w, u \otimes x \rangle &= \left\langle \sum_{j=1}^{m} v_j e_j \otimes \sum_{k=1}^{n} w_k f_k,\ \sum_{j=1}^{m} u_j e_j \otimes \sum_{k=1}^{n} x_k f_k \right\rangle \\
&= \left\langle \sum_{k=1}^{n} \sum_{j=1}^{m} v_j w_k\, e_j \otimes f_k,\ \sum_{k=1}^{n} \sum_{j=1}^{m} u_j x_k\, e_j \otimes f_k \right\rangle \\
&= \sum_{k=1}^{n} \sum_{j=1}^{m} v_j \overline{u_j} w_k \overline{x_k} \\
&= \left(\sum_{j=1}^{m} v_j \overline{u_j} \right) \left(\sum_{k=1}^{n} w_k \overline{x_k} \right) \\
&= \langle v, u \rangle \langle w, x \rangle.
\end{aligned}$$

There is only one inner product on $V \otimes W$ such that $\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle \langle w, x \rangle$ for all $v, u \in V$ and $w, x \in W$ because every element of $V \otimes W$ can be written as a linear combination of elements of the form $v \otimes w$ [by 9.74(b)]. □

The definition below of a natural inner product on $V \otimes W$ is now justified by 9.80. We could not have simply defined $\langle v \otimes w, u \otimes x\rangle$ to be $\langle v, u\rangle\langle w, x\rangle$ (and then used additivity in each slot separately to extend the definition to $V \otimes W$) without some proof because elements of $V \otimes W$ do not have unique representations as finite sums of elements of the form $v \otimes w$.

Definition 9.82

Inner product on tensor product of two inner product spaces

Suppose V and W are inner product spaces. The inner product on $V \otimes W$ is the unique function $\langle \cdot, \cdot \rangle$ from $(V \otimes W) \times (V \otimes W)$ to $\mathbb{F}$ such that

$$\langle v \otimes w, u \otimes x\rangle = \langle v, u\rangle\langle w, x\rangle$$

for all $v, u \in V$ and $w, x \in W$.

Take $u = v$ and $x = w$ in the equation above and then take square roots to show that

$$\|v \otimes w\| = \|v\| \, \|w\|$$

for all $v \in V$ and all $w \in W$.

The construction of the inner product in the proof of 9.80 depended on an orthonormal basis $e_1, \dots, e_m$ of V and an orthonormal basis $f_1, \dots, f_n$ of W. Formula 9.81 implies that $\{e_j \otimes f_k\}_{j=1,\dots,m;\, k=1,\dots,n}$ is a doubly indexed orthonormal list in $V \otimes W$ and hence is an orthonormal basis of $V \otimes W$ [by 9.74(b)]. The importance of the next result arises because the orthonormal bases used there can be different from the orthonormal bases used to define the inner product in 9.80. Although the notation for the bases is the same in the proof of 9.80 and in the result below, think of them as two different sets of orthonormal bases.

Result 9.83

Orthonormal basis of $V \otimes W$

Suppose V and W are inner product spaces, and $e_1, \dots, e_m$ is an orthonormal basis of V and $f_1, \dots, f_n$ is an orthonormal basis of W. Then

$$\{e_j \otimes f_k\}_{j=1,\dots,m;\, k=1,\dots,n}$$

is an orthonormal basis of $V \otimes W$.

Proof. We know that $\{e_j \otimes f_k\}_{j=1,\dots,m;\, k=1,\dots,n}$ is a basis of $V \otimes W$ [by 9.74(b)]. Thus we only need to verify orthonormality. To do this, suppose $j, M \in \{1, \dots, m\}$ and $k, N \in \{1, \dots, n\}$. Then

$$\langle e_j \otimes f_k, e_M \otimes f_N\rangle = \langle e_j, e_M\rangle\langle f_k, f_N\rangle = \begin{cases} 1 & \text{if } j = M \text{ and } k = N, \\ 0 & \text{otherwise.} \end{cases}$$

Hence the doubly indexed list $\{e_j \otimes f_k\}_{j=1,\dots,m;\, k=1,\dots,n}$ is indeed an orthonormal basis of $V \otimes W$. □

See Exercise 11 for an example of how the inner product structure on $V \otimes W$ interacts with operators on V and W.

Tensor Product of Multiple Vector Spaces

We have been discussing properties of the tensor product of two finite-dimensional vector spaces. Now we turn our attention to the tensor product of multiple finite-dimensional vector spaces. This generalization requires no new ideas, only some slightly more complicated notation. Readers with a good understanding of the tensor product of two vector spaces should be able to make the extension to the tensor product of more than two vector spaces.

Thus in this subsection, no proofs will be provided. The definitions and the statements of results that will be provided should be enough information to enable readers to fill in the details, using what has already been learned about the tensor product of two vector spaces.

We begin with the following notational assumption.

Notation 9.84

$V_1, \ldots, V_m$

For the rest of this subsection, m denotes an integer greater than 1 and $V_1, \ldots, V_m$ denote finite-dimensional vector spaces.

The notion of an m-linear functional, which we are about to define, generalizes the notion of a bilinear functional (see 9.68). Recall that the use of the word "functional" indicates that we are mapping into the scalar field $\mathbb{F}$. Recall also that the terminology "m-linear form" is used in the special case $V_1 = \cdots = V_m$ (see 9.25). The notation $\mathcal{B}(V_1, \ldots, V_m)$ generalizes our previous notation $\mathcal{B}(V, W)$.

Definition 9.85

m-linear functional, the vector space $\mathcal{B}(V_1, \ldots, V_m)$

- An *m-linear functional* on $V_1 \times \cdots \times V_m$ is a function $\beta : V_1 \times \cdots \times V_m \to \mathbb{F}$ that is a linear functional in each slot when the other slots are held fixed.
- The vector space of m-linear functionals on $V_1 \times \cdots \times V_m$ is denoted by $\mathcal{B}(V_1, \ldots, V_m)$.

Example 9.86 **m-linear functional**

Suppose $\varphi_k \in V_k'$ for each $k \in \{1, \ldots, m\}$. Define $\beta : V_1 \times \cdots \times V_m \to \mathbb{F}$ by

$$\beta(v_1, \ldots, v_m) = \varphi_1(v_1) \times \cdots \times \varphi_m(v_m).$$

Then β is an m-linear functional on $V_1 \times \cdots \times V_m$.

The next result can be proved by imitating the proof of 9.70.

Dimension of the vector space of m-linear functionals — Result 9.87

$$\dim \mathcal{B}(V_1, \dots, V_m) = (\dim V_1) \times \cdots \times (\dim V_m).$$

Now we can define the tensor product of multiple vector spaces and the tensor product of elements of those vector spaces. The following definition is completely analogous to our previous definition (9.71) in the case $m = 2$.

Tensor product, $V_1 \otimes \cdots \otimes V_m$, $v_1 \otimes \cdots \otimes v_m$ — Definition 9.88

- The *tensor product* $V_1 \otimes \cdots \otimes V_m$ is defined to be $\mathcal{B}(V_1', \dots, V_m')$.
- For $v_1 \in V_1, \dots, v_m \in V_m$, the *tensor product* $v_1 \otimes \cdots \otimes v_m$ is the element of $V_1 \otimes \cdots \otimes V_m$ defined by

$$(v_1 \otimes \cdots \otimes v_m)(\varphi_1, \dots, \varphi_m) = \varphi_1(v_1) \cdots \varphi_m(v_m)$$

for all $(\varphi_1, \dots, \varphi_m) \in V_1' \times \cdots \times V_m'$.

The next result can be proved by following the pattern of the proof of the analogous result when $m = 2$ (see 9.72).

Dimension of the tensor product — Result 9.89

$$\dim(V_1 \otimes \cdots \otimes V_m) = (\dim V_1) \cdots (\dim V_m).$$

Our next result generalizes 9.74.

Basis of $V_1 \otimes \cdots \otimes V_m$ — Result 9.90

Suppose $\dim V_k = n_k$ and $e_1^k, \dots, e_{n_k}^k$ is a basis of V_k for $k = 1, \dots, m$. Then

$$\{e_{j_1}^1 \otimes \cdots \otimes e_{j_m}^m\}_{j_1 = 1, \dots, n_1; \cdots; j_m = 1, \dots, n_m}$$

is a basis of $V_1 \otimes \cdots \otimes V_m$.

Suppose $m = 2$ and $e_1^1, \dots, e_{n_1}^1$ is a basis of V_1 and $e_1^2, \dots, e_{n_2}^2$ is a basis of V_2. Then with respect to the basis $\{e_{j_1}^1 \otimes e_{j_2}^2\}_{j_1 = 1, \dots, n_1;\, j_2 = 1, \dots, n_2}$ in the result above, the coefficients of an element of $V_1 \otimes V_2$ can be represented by an n_1-by-n_2 matrix that contains the coefficient of $e_{j_1}^1 \otimes e_{j_2}^2$ in row j_1, column j_2. Thus we need a matrix, which is an array specified by two indices, to represent an element of $V_1 \otimes V_2$.

If $m > 2$, then the result above shows that we need an array specified by m indices to represent an arbitrary element of $V_1 \otimes \cdots \otimes V_m$. Thus tensor products may appear when we deal with objects specified by arrays with multiple indices.

The next definition generalizes the notion of a bilinear map (see 9.77). As with bilinear maps, the target space can be an arbitrary vector space.

m-linear map Definition 9.91

An *m-linear map* from $V_1 \times \cdots \times V_m$ to a vector space U is a function $\Gamma: V_1 \times \cdots \times V_m \to U$ that is a linear map in each slot when the other slots are held fixed.

The next result can be proved by following the pattern of the proof of 9.79.

Converting m-linear maps to linear maps Result 9.92

Suppose U is a vector space.

(a) Suppose that $\Gamma: V_1 \times \cdots \times V_m \to U$ is an m-linear map. Then there exists a unique linear map $\hat{\Gamma}: V_1 \otimes \cdots \otimes V_m \to U$ such that

$$\hat{\Gamma}(v_1 \otimes \cdots \otimes v_m) = \Gamma(v_1, \ldots, v_m)$$

for all $(v_1, \ldots, v_m) \in V_1 \times \cdots \times V_m$.

(b) Conversely, suppose $T: V_1 \otimes \cdots \otimes V_m \to U$ is a linear map. Then there exists a unique m-linear map $T^{\#}: V_1 \times \cdots \times V_m \to U$ such that

$$T^{\#}(v_1, \ldots, v_m) = T(v_1 \otimes \cdots \otimes v_m)$$

for all $(v_1, \ldots, v_m) \in V_1 \times \cdots \times V_m$.

See Exercises 12 and 13 for tensor products of multiple inner product spaces.

Exercises

9D.1 Suppose $v \in V$ and $w \in W$. Prove that $v \otimes w = 0$ if and only if $v = 0$ or $w = 0$.

9D.2 Give an example of six distinct vectors $v_1, v_2, v_3, w_1, w_2, w_3$ in $\mathbb{R}^3$ such that

$$v_1 \otimes w_1 + v_2 \otimes w_2 + v_3 \otimes w_3 = 0$$

but none of $v_1 \otimes w_1, v_2 \otimes w_2, v_3 \otimes w_3$ is a scalar multiple of another element of this list.

9D.3 Suppose that $v_1, \ldots, v_m$ is a linearly independent list in V. Suppose also that $w_1, \ldots, w_m$ is a list in W such that

$$v_1 \otimes w_1 + \cdots + v_m \otimes w_m = 0.$$

Prove that $w_1 = \cdots = w_m = 0$.

9D.4 Suppose $\dim V > 1$ and $\dim W > 1$. Prove that

$$\{v \otimes w : (v, w) \in V \times W\}$$

is not a subspace of $V \otimes W$.

This exercise implies that if $\dim V > 1$ *and* $\dim W > 1$, *then*

$$\{v \otimes w : (v, w) \in V \times W\} \neq V \otimes W.$$

9D.5 Suppose m and n are positive integers. For $v \in \mathbb{F}^m$ and $w \in \mathbb{F}^n$, identify $v \otimes w$ with an m-by-n matrix as in Example 9.76. With that identification, show that the set

$$\{v \otimes w : v \in \mathbb{F}^m \text{ and } w \in \mathbb{F}^n\}$$

is the set of m-by-n matrices (with entries in $\mathbb{F}$) that have rank at most one.

9D.6 Suppose m and n are positive integers. Give a description, analogous to Exercise 5, of the set of m-by-n matrices (with entries in $\mathbb{F}$) that have rank at most two.

9D.7 Suppose $\dim V > 2$ and $\dim W > 2$. Prove that

$$\{v_1 \otimes w_1 + v_2 \otimes w_2 : v_1, v_2 \in V \text{ and } w_1, w_2 \in W\} \neq V \otimes W.$$

9D.8 Suppose $v_1, \dots, v_m \in V$ and $w_1, \dots, w_m \in W$ are such that

$$v_1 \otimes w_1 + \cdots + v_m \otimes w_m = 0.$$

Suppose that U is a vector space and $\Gamma\colon V \times W \to U$ is a bilinear map. Show that

$$\Gamma(v_1, w_1) + \cdots + \Gamma(v_m, w_m) = 0.$$

9D.9 Suppose $S \in \text{End}(V)$ and $T \in \text{End}(W)$. Prove that there exists a unique operator on $V \otimes W$ that takes $v \otimes w$ to $Sv \otimes Tw$ for all $v \in V$ and $w \in W$.

In an abuse of notation, the operator on $V \otimes W$ *given by this exercise is often called* $S \otimes T$.

9D.10 Suppose $S \in \text{End}(V)$ and $T \in \text{End}(W)$. Prove that $S \otimes T$ is an invertible operator on $V \otimes W$ if and only if both S and T are invertible operators. Also, prove that if both S and T are invertible operators, then $(S \otimes T)^{-1} = S^{-1} \otimes T^{-1}$, where we are using the notation from the comment after Exercise 9.

9D.11 Suppose V and W are inner product spaces. Prove that if $S \in \text{End}(V)$ and $T \in \text{End}(W)$, then $(S \otimes T)^* = S^* \otimes T^*$, where we are using the notation from the comment after Exercise 9.

9D.12 Suppose that $V_1, \ldots, V_m$ are finite-dimensional inner product spaces. Prove that there is a unique inner product on $V_1 \otimes \cdots \otimes V_m$ such that

$$\langle v_1 \otimes \cdots \otimes v_m, u_1 \otimes \cdots \otimes u_m \rangle = \langle v_1, u_1 \rangle \cdots \langle v_m, u_m \rangle$$

for all $(v_1, \ldots, v_m)$ and $(u_1, \ldots, u_m)$ in $V_1 \times \cdots \times V_m$.

Note that the equation above implies that

$$\|v_1 \otimes \cdots \otimes v_m\| = \|v_1\| \times \cdots \times \|v_m\|$$

for all $(v_1, \ldots, v_m) \in V_1 \times \cdots \times V_m$.

9D.13 Suppose that $V_1, \ldots, V_m$ are finite-dimensional inner product spaces and $V_1 \otimes \cdots \otimes V_m$ is made into an inner product space using the inner product from Exercise 12. Suppose $e_1^k, \ldots, e_{n_k}^k$ is an orthonormal basis of V_k for each $k = 1, \ldots, m$. Show that the list

$$\{e_{j_1}^1 \otimes \cdots \otimes e_{j_m}^m\}_{j_1 = 1, \ldots, n_1; \cdots; j_m = 1, \ldots, n_m}$$

is an orthonormal basis of $V_1 \otimes \cdots \otimes V_m$.

Solutions

Solutions to Axler's Linear Algebra Done Right, 4th Edition, by blanketism. The manual was begun on July 4, 2026 and completed on November 2, 2026; the version used here was last edited on November 5, 2026.

Preface

The solutions printed here are those of the manual *Solutions to Axler's Linear Algebra Done Right, 4th Edition* by blanketism, which accompanies the fourth edition of Sheldon Axler's text. They are reproduced with the author's solutions unchanged in content; only the typesetting and the notation have been adapted to this book. In particular, $\mathcal{L}(V, W)$ and $\mathcal{L}(V)$ are written $\operatorname{Hom}(V, W)$ and $\operatorname{End}(V)$, the polynomials $\mathcal{P}_m(\mathbb{F})$ are written $\mathbb{F}[X]_{\leq m}$, null spaces and ranges are written with ker and im, and the matrix of an operator is written $\operatorname{Mat}(T)$.

The manual is meant as a companion to the text and not as a replacement for reading it and attempting the exercises independently. Its author recommends trying each exercise before consulting the solution, since linear algebra, and particularly Axler's determinant-free and operator-theoretic approach, is best learned through active engagement with the material. The manual is written primarily for students who meet rigorous linear algebra for the first time, and it aims at clarity, accessibility and completeness.

The author does not guarantee that the solutions are free of errors, and welcomes corrections, suggestions and feedback through the following channels:

- Discord Errata Server: `https://discord.gg/qecQk37vfm`
- Reddit: `https://www.reddit.com/user/ansv9a8fdh3/`
- GitHub: `https://github.com/blanketism/Axler-Solutions`

In this edition

The solutions cover all nine chapters. They are printed below, grouped by section. The manual's restatements of the exercises are omitted, since the exercises appear in the text. The alternate proofs of the manual's appendix, which the manual marks with a star at the exercise, follow the solution of that exercise, and the solution says so. A short note at the head of a solution is one of the following:

- **Axler Hint.** *A hint of the author of the text that the book does not print with the exercise.*
- **Axler Note.** *A note of the author of the text that the book does not print with the exercise.*
- **Manual Remark.** *A remark of the manual's author, offering clarification or a different perspective on the exercise.*

Chapter 1

§1A

Solution 1A.1 Let $\alpha = a + bi$ and $\beta = c + di$. Then

$$\begin{aligned}\alpha + \beta &= (a + bi) + (c + di)\\ &= (a + c) + (b + d)i\\ &= (c + a) + (d + b)i\\ &= (c + di) + (a + bi) = \beta + \alpha.\end{aligned}$$ □

Solution 1A.2 Let $\alpha = a + bi$, $\beta = c + di$, and $\gamma = e + fi$. Then

$$\begin{aligned}(\alpha + \beta) + \gamma &= ((a + bi) + (c + di)) + (e + fi)\\ &= ((a + c) + (b + d)i) + (e + fi)\\ &= (a + c + e) + (b + d + f)i\\ &= (a + (c + e)) + (b + (d + f))i\\ &= (a + bi) + ((c + di) + (e + fi)) = \alpha + (\beta + \gamma).\end{aligned}$$ □

Solution 1A.3 Let $\alpha = a + bi$, $\beta = c + di$, and $\gamma = e + fi$. Then

$$\begin{aligned}(\alpha\beta)\gamma &= ((a + bi)(c + di))(e + fi)\\ &= ((ac - bd) + (ad + bc)i)(e + fi)\\ &= (ace - bde - adf - bcf) + (acf - bdf + ade + bce)i\\ &= (ace - adf - bde - bcf) + (acf + ade + bce - bdf)i\\ &= (a + bi)((ce - df) + (cf + de)i)\\ &= (a + bi)((c + di)(e + fi)) = \alpha(\beta\gamma).\end{aligned}$$ □

Solution 1A.4 Let $\alpha = a + bi$, $\beta = c + di$, and $\gamma = e + fi$. Then

$$\begin{aligned}\gamma(\alpha + \beta) &= (e + fi)((a + bi) + (c + di))\\ &= (e + fi)((a + c) + (b + d)i)\\ &= (e(a + c) - f(b + d)) + (e(b + d) + f(a + c))i\\ &= (ea - fb + ec - fd) + (eb + fa + ed + fc)i\\ &= (ea - fb) + (eb + fa)i + (ec - fd) + (ed + fc)i\\ &= (e + fi)(a + bi) + (e + fi)(c + di) = \gamma\alpha + \gamma\beta.\end{aligned}$$ □

Solution 1A.5 Let $\alpha = a + bi$. We want to find $\beta = c + di$ such that $\alpha + \beta = 0$. This gives

$$(a + bi) + (c + di) = (a + c) + (b + d)i = 0 + 0i.$$

Solving for c and d, we get $c = -a$ and $d = -b$. Hence, $\beta = c + di = -a - bi = -\alpha$. To see that β is unique, suppose there is another $\gamma \in \mathbb{C}$ such that $\alpha + \gamma = 0$. Then

$$\beta = \beta + 0 = \beta + (\alpha + \gamma) = (\beta + \alpha) + \gamma = 0 + \gamma = \gamma.$$

Hence, β is unique. □

Solution 1A.6 Let $\alpha = a + bi$ such that $\alpha \neq 0$. Then at least one of a or b is nonzero. We want to find $\beta = c + di$ such that $\alpha\beta = 1$. Then

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i = 1 + 0i$$

which gives the system of equations

$$\begin{cases} ac - bd = 1 \\ ad + bc = 0 \end{cases} \implies \begin{cases} a^2c - abd = a \\ b^2c + abd = 0. \end{cases}$$

Adding the two equations, we get

$$(a^2+b^2)c = a \implies c = \frac{a}{a^2 + b^2} \implies abd + b^2\left(\frac{a}{a^2 + b^2}\right) = 0 \implies d = \frac{-b}{a^2 + b^2}.$$

Hence,

$$\beta = c + di = \frac{a}{a^2 + b^2} + \frac{-b}{a^2 + b^2}i.$$

If $\gamma \in \mathbb{C}$ is another complex number such that $\alpha\gamma = 1$, then

$$\gamma = \gamma 1 = \gamma(\alpha\beta) = (\gamma\alpha)\beta = 1\beta = \beta.$$

Hence, β is unique. □

Solution 1A.7 We have

$$\begin{aligned} (-1 + \sqrt{3}i)^3 &= (-1)^3 + 3(-1)^2(\sqrt{3}i) + 3(-1)(\sqrt{3}i)^2 + (\sqrt{3}i)^3 \\ &= -1 + 3\sqrt{3}i - 9i^2 - 3\sqrt{3}i^3 \\ &= -1 + 3\sqrt{3}i + 9 - 3\sqrt{3}i = 8. \end{aligned}$$

Hence,

$$\left(\frac{-1 + \sqrt{3}i}{2}\right)^3 = \frac{(-1 + \sqrt{3}i)^3}{2^3} = \frac{8}{8} = 1.$$ □

Solution 1A.8 Let $z = a + bi \in \mathbb{C}$ such that $z^2 = i$. Then

$$(a + bi)^2 = a^2 - b^2 + 2abi = 0 + 1i,$$

giving the system of equations

$$\begin{cases} a^2 - b^2 = 0 \\ 2ab = 1. \end{cases}$$

The first equation yields $a = \pm b$. If $a = b$, then

$$2b^2 = 1 \implies b^2 = \frac{1}{2} \implies b = \pm\frac{1}{\sqrt{2}} \implies a = \pm\frac{1}{\sqrt{2}}.$$

If $a = -b$, then $-2b^2 = 1$, which has no real solutions. Hence, the two distinct square roots of i are

$$\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}i \quad \text{and} \quad -\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}i. \qquad \square$$

Solution 1A.9 Let $x = (x_i)_{i=1}^n \in \mathbb{R}^4$. The equation becomes

$$(4, -3, 1, 7) + 2(x_1, x_2, x_3, x_4) = (5, 9, -6, 8),$$

which simplifies to the system of equations

$$\begin{cases} 5 = 4 + 2x_1 \\ 9 = -3 + 2x_2 \\ -6 = 1 + 2x_3 \\ 8 = 7 + 2x_4 \end{cases} \implies \begin{cases} 2x_1 = 1 \\ 2x_2 = 12 \\ 2x_3 = -7 \\ 2x_4 = 1 \end{cases} \implies x = \left(\frac{1}{2}, 6, -\frac{7}{2}, \frac{1}{2}\right). \qquad \square$$

Solution 1A.10 If such a λ exists, then we have the system of equations

$$\begin{cases} \lambda(2 - 3i) = 12 - 5i \\ \lambda(5 + 4i) = 7 + 22i \\ \lambda(-6 + 7i) = -32 - 9i \end{cases}$$

The first equation gives

$$\lambda = \frac{12 - 5i}{2 - 3i} = \frac{(12 - 5i)(2 + 3i)}{(2 - 3i)(2 + 3i)} = \frac{24 + 36i - 10i - 15i^2}{4 + 9} = \frac{24 + 26i + 15}{13} = 3+2i.$$

However, using this value for λ in the third equation, we get

$$(3+2i)(-6+7i) = -18+21i-12i+14i^2 = -18+9i-14 = -32+9i \neq -32-9i. \ \square$$

Solution 1A.11 Let $x = (x_i)_{i=1}^n, y = (y_i)_{i=1}^n, z = (z_i)_{i=1}^n \in \mathbb{F}^n$. Then

$$\begin{aligned}(x+y)+z &= ((x_i)_1^n + (y_i)_1^n) + (z_i)_1^n \\ &= (x_i + y_i)_1^n + (z_i)_1^n \\ &= ((x_i + y_i) + z_i)_1^n \\ &= (x_i + (y_i + z_i))_1^n \\ &= (x_i)_1^n + (y_i + z_i)_1^n \\ &= (x_i)_1^n + ((y_i)_1^n + (z_i)_1^n) = x + (y + z).\end{aligned}$$ □

Solution 1A.12 Let $a, b \in \mathbb{F}$ and $x = (x_i)_{i=1}^n \in \mathbb{F}^n$. Then

$$(ab)x = (ab)(x_i)_1^n = ((ab)x_i)_1^n = (a(bx)_i)_1^n = a(bx_i)_1^n = a(b(x_i)_1^n) = a(bx).$$ □

Solution 1A.13 Let $x = (x_i)_{i=1}^n \in \mathbb{F}^n$. Then

$$1x = 1(x_i)_1^n = (1x_i)_1^n = (x_i)_1^n = x.$$ □

Solution 1A.14 Let $\lambda \in \mathbb{F}$ and $x = (x_i)_{i=1}^n, y = (y_i)_{i=1}^n \in \mathbb{F}^n$. Then

$$\begin{aligned}\lambda(x+y) &= \lambda((x_i)_1^n + (y_i)_1^n) \\ &= \lambda(x_i + y_i)_1^n \\ &= (\lambda(x_i + y)_i)_1^n \\ &= (\lambda x_i + \lambda y_i)_1^n \\ &= (\lambda x_i)_1^n + (\lambda y_i)_1^n \\ &= \lambda(x_i)_1^n + \lambda(y_i)_1^n = \lambda x + \lambda y.\end{aligned}$$ □

Solution 1A.15 Let $a, b \in \mathbb{F}$ and $x = (x_i)_{i=1}^n \in \mathbb{F}^n$. Then

$$(a+b)x = (a+b)(x_i)_1^n = ((a+b)x_i)_1^n = (ax+bx_i)_1^n = (ax_i)_1^n + (bx_i)_1^n = a(x_i)_1^n + b(x_i)_1^n = ax+bx.$$ □

§1B

Solution 1B.1 Let $v \in V$. By the definition of additive inverse, we have

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

Since additive inverses are unique, we conclude that $-(-v) = v$. □

Solution 1B.2 If $a = 0$, then we are done. Suppose $a \neq 0$. Then

$$v = 1v = \frac{1}{a}(av) = \frac{1}{a} \cdot 0 = 0. \qquad \square$$

Solution 1B.3

Manual Remark. *The solution says "dividing by 3" in the sense that we're dividing by the scalar 3 in the field* $\mathbb{F}$*, not a vector in* V*.*

Solving for x explicitly, we obtain

$$3x = w - v \implies x = \frac{1}{3}(w - v).$$

This shows that there exists $x \in V$ such that $v + 3x = w$. To see that x is unique, suppose there exists another $y \in V$ such that $v + 3y = w$. Then dividing by 3, we get

$$3x = w - v = 3y \implies x = y. \qquad \square$$

Solution 1B.4 The empty set fails to satisfy the requirement that there exists a zero vector $0 \in V$ such that $v + 0 = v$ for all $v \in V$ since it contains no elements. $\square$

Solution 1B.5 $\Rightarrow$ Suppose that $0v = 0$ for all $v \in V$. Let $v \in V$. Then

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

Hence, $(-1)v$ is the additive inverse of v.
$\Leftarrow$ Conversely, suppose that for every $v \in V$, there exists an additive inverse $-v \in V$ such that $v + (-v) = 0$. Then $0v = (0 + 0)v = 0v + 0v$. Adding $-0v$ to both sides, we get $0 = 0v$. Hence, $0v = 0$ for all $v \in V$. $\square$

Solution 1B.6 No, $\mathbb{R} \cup \{\infty, -\infty\}$ is not a vector space over $\mathbb{R}$ as it does not satisfy the associative property of addition. Let $t \in \mathbb{R} \setminus \{0\}$. Then

$$(t + \infty) + (-\infty) = \infty + (-\infty) = 0 \neq t = t + 0 = t + (\infty + (-\infty)). \qquad \square$$

Solution 1B.7 Let $f, g \in V^S$ and $t \in \mathbb{F}$. Define addition and scalar multiplication as follows:

$$(f + g)(s) = f(s) + g(s), \quad (tf)(s) = t(f(s))$$

for all $s \in S$. We will show that V^S is a vector space with these operations.

1. **Commutativity:** Let $f, g \in V^S$. Then for all $s \in S$,

$$(f + g)(s) = f(s) + g(s) = g(s) + f(s) = (g + f)(s).$$

2. **Associativity:** Let $f, g, h \in V^S$ and $t, u \in \mathbb{F}$. Then for all $s \in S$,

$$\begin{aligned}((f+g)+h)(s) &= (f+g)(s) + h(s)\\ &= (f(s)+g(s)) + h(s)\\ &= f(s) + (g(s)+h(s))\\ &= f(s) + (g+h)(s)\\ &= (f+(g+h))(s),\end{aligned}$$

and

$$\begin{aligned}((tu)f)(s) &= (tu)(f(s))\\ &= t(u(f(s)))\\ &= t(uf)(s)\\ &= (t(uf))(s).\end{aligned}$$

3. **Additive identity:** Let $0 \in V$ be the additive identity in V. Define the zero function $0_S \in V^S$ by $0_S(s) = 0$ for all $s \in S$. Then for all $f \in V^S$ and $s \in S$,

$$(f + 0_S)(s) = f(s) + 0_S(s) = f(s) + 0 = f(s) = 0 + f(s) = (0_S + f)(s).$$

Hence, 0_S is the additive identity in V^S.

4. **Additive inverse:** Let $f \in V^S$. Define the function $-f \in V^S$ by $(-f)(s) = -f(s)$ for all $s \in S$. Then for all $s \in S$,

$$(f + (-f))(s) = f(s) + (-f)(s) = f(s) + (-f(s)) = 0 = 0_S(s).$$

Hence, $-f$ is the additive inverse of f in V^S.

5. **Multiplicative identity:** For all $f \in V^S$ and $s \in S$,

$$(1f)(s) = 1(f(s)) = f(s).$$

6. **Distributive properties:** Let $f, g \in V^S$ and $t, u \in \mathbb{F}$. Then for all $s \in S$,

$$\begin{aligned}((t+u)f)(s) &= (t+u)(f(s))\\ &= t(f(s)) + u(f(s))\\ &= (tf)(s) + (uf)(s)\\ &= (tf+uf)(s),\end{aligned}$$

and

$$\begin{aligned}(t(f+g))(s) &= t((f+g)(s))\\ &= t(f(s)+g(s))\\ &= t(f(s)) + t(g(s))\\ &= (tf)(s) + (tg)(s)\\ &= (tf+tg)(s).\end{aligned}$$

□

Solution 1B.8 We prove that $V_{\mathbb{C}}$ is a complex vector space by verifying the vector space axioms.

1. **Commutativity:** Let $u_1 + v_1 i, u_2 + v_2 i \in V_{\mathbb{C}}$. Then

$$(u_1+v_1 i)+(u_2+v_2 i) = (u_1+u_2)+i(v_1+v_2) = (u_2+u_1)+i(v_2+v_1) = (u_2+v_2 i)+(u_1+v_1 i).$$

2. **Associativity:** Let $u_1 + v_1 i, u_2 + v_2 i, u_3 + v_3 i \in V_{\mathbb{C}}$. Then

$$\begin{aligned}((u_1 + v_1 i) + (u_2 + v_2 i)) + (u_3 + v_3 i) &= ((u_1 + u_2) + i(v_1 + v_2)) + (u_3 + v_3 i)\\ &= ((u_1 + u_2) + u_3) + i((v_1 + v_2) + v_3)\\ &= (u_1 + (u_2 + u_3)) + i(v_1 + (v_2 + v_3))\\ &= (u_1 + v_1 i) + ((u_2 + v_2 i) + (u_3 + v_3 i)).\end{aligned}$$

   Similarly, let $a + bi, c + di \in \mathbb{C}$ and $u + vi \in V_{\mathbb{C}}$. Then

$$\begin{aligned}((a + bi)(c + di))(u + vi) &= ((ac - bd) + i(ad + bc))(u + vi)\\ &= ((ac - bd)u - (ad + bc)v) + i((ac - bd)v + (ad + bc)u)\\ &= (a(cu - dv) - b(du + cv)) + i(a(du + cv) + b(cu - dv))\\ &= (a + bi)((cu - dv) + i(du + cv))\\ &= (a + bi)((c + di)(u + vi)).\end{aligned}$$

3. **Additive identity:** Let $0 \in V$ be the additive identity in V. Then for all $u + vi \in V_{\mathbb{C}}$,

$$(u + vi) + (0 + 0i) = (u + 0) + i(v + 0) = u + vi = (0 + 0i) + (u + vi).$$

   Hence, $0 + 0i$ is the additive identity in $V_{\mathbb{C}}$.

4. **Additive inverse:** Let $u + vi \in V_{\mathbb{C}}$. Then

$$(u + vi) + (-u + (-v)i) = (u - u) + i(v - v) = 0 + 0i.$$

   Hence, $-u + (-v)i$ is the additive inverse of $u + vi$ in $V_{\mathbb{C}}$.

5. **Multiplicative identity:** For all $u + vi \in V_{\mathbb{C}}$,

$$(1 + 0i)(u + vi) = (1u - 0v) + i(1v + 0u) = u + vi.$$

6. **Distributive properties:** Let $a+bi, c+di \in \mathbb{C}$ and $u_1+v_1 i, u_2+v_2 i \in V_{\mathbb{C}}$. Then

$$\begin{aligned}(a + bi)((u_1 + v_1 i) + (u_2 + v_2 i)) &= (a + bi)((u_1 + u_2) + i(v_1 + v_2))\\ &= (a(u_1 + u_2) - b(v_1 + v_2)) + i(a(v_1 + v_2) + b(u_1 + u_2))\\ &= (au_1 - bv_1) + i(av_1 + bu_1) + (au_2 - bv_2) + i(av_2 + bu_2)\\ &= (a + bi)(u_1 + v_1 i) + (a + bi)(u_2 + v_2 i),\end{aligned}$$

   and

$$\begin{aligned}((a + bi) + (c + di))(u_1 + v_1 i) &= ((a + c) + (b + d)i)(u_1 + v_1 i)\\ &= ((a + c)u_1 - (b + d)v_1) + i((a + c)v_1 + (b + d)u_1)\\ &= (au_1 - bv_1) + i(av_1 + bu_1) + (cu_1 - dv_1) + i(cv_1 + du_1)\\ &= (a + bi)(u_1 + v_1 i) + (c + di)(u_1 + v_1 i). \qquad \square\end{aligned}$$

§1C

Solution 1C.1 Let U be the subset of $\mathbb{F}^3$ in question.

(a) Since $0 + 2(0) + 3(0) = 0$, $(0,0,0) \in U$, so U is nonempty. Let $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3) \in U$ and $\alpha \in \mathbb{F}$. Then

$$(u_1+v_1)+2(u_2+v_2)+3(u_3+v_3) = (u_1+2u_2+3u_3)+(v_1+2v_2+3v_3) = 0+0 = 0,$$

and

$$\alpha u_1 + 2(\alpha u_2) + 3(\alpha u_3) = \alpha(u_1 + 2u_2 + 3u_3) = \alpha(0) = 0.$$

Hence, $u + v \in U$ and $\alpha u \in U$, so U is a subspace of $\mathbb{F}^3$.

(b) Since $0+2(0)+3(0) = 0 \neq 4$, $(0,0,0) \notin U$, so U does not contain the zero vector. Therefore, U is not a subspace of $\mathbb{F}^3$.

(c) Since $1 \cdot 0 \cdot 0 = 0$ and $0 \cdot 1 \cdot 1 = 0$, $(1,0,0), (0,1,1) \in U$. However, $(1,0,0) + (0,1,1) = (1,1,1)$ and $1 \cdot 1 \cdot 1 = 1 \neq 0$, so $(1,1,1) \notin U$. Hence, U is not closed under addition, so U is not a subspace of $\mathbb{F}^3$.

(d) Since $5(0) = 0$, $(0,0,0) \in U$, so U is nonempty. Let $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3) \in U$ and $\alpha \in \mathbb{F}$. Then

$$u_1 + v_1 = 5u_3 + 5v_3 = 5(u_3 + v_3),$$

and

$$\alpha u_1 = \alpha(5u_3) = 5(\alpha u_3).$$

Hence, $u + v \in U$ and $\alpha u \in U$, so U is a subspace of $\mathbb{F}^3$. □

Solution 1C.2 Let U be the subsets in question.

(a) $\Rightarrow$ Suppose U is a subspace of $\mathbb{F}^4$. Then it must contain the zero vector $(0,0,0,0)$. Then $0 = 5(0) + b$, which implies $b = 0$.

$\Leftarrow$ Suppose $b = 0$. Then $U = \{(x_1, x_2, x_3, x_4) \in \mathbb{F}^4 : x_3 = 5x_4\}$. The proof is similar to the proof in Exercise 1C1(d), so we conclude that U is a subspace of $\mathbb{F}^4$.

(b) Since the 0 function is continuous, $0 \in U$. Let $f, g \in U$ and $\alpha \in \mathbb{R}$. Since both $f + g$ and αf are continuous over $[0,1]$, $f + g, \alpha f \in U$. Therefore, U is a subspace of $\mathbb{R}^{[0,1]}$.

(c) Since the 0 function is differentiable on $\mathbb{R}$, $0 \in U$. Let $f, g \in U$ and $\alpha \in \mathbb{R}$. Since both $f+g$ and αf are differentiable on $\mathbb{R}$, $f+g, \alpha f \in U$. Hence, U is a subspace of $\mathbb{R}^{\mathbb{R}}$.

(d) $\Rightarrow$ Suppose U is a subspace of $\mathbb{R}^{(0,3)}$. Then $0 \in U$, so $0'(2) = 0 = b$, which implies $b = 0$.

$\Leftarrow$ Suppose $b = 0$. Then

$$U = \{f \in \mathbb{R}^{(0,3)} : f'(2) = 0\}.$$

The zero function is in U since $0'(2) = 0$. Let $f, g \in U$ and $\alpha \in \mathbb{R}$. Then

$$(f + g)'(2) = f'(2) + g'(2) = 0 + 0 = 0,$$

and

$$(\alpha f)'(2) = \alpha f'(2) = \alpha \cdot 0 = 0.$$

Thus, $f + g, \alpha f \in U$, so U is a subspace of $\mathbb{R}^{(0,3)}$.

(e) The sequence of zeroes has limit 0, so it is in U. Let $\{z_n\}_{n=1}^{\infty}, \{w_n\}_{n=1}^{\infty} \in U$ and $\alpha \in \mathbb{C}$. Then

$$\lim_{n\to\infty}(z_n + w_n) = \lim_{n\to\infty} z_n + \lim_{n\to\infty} w_n = 0 + 0 = 0,$$

and

$$\lim_{n\to\infty}(\alpha z_n) = \alpha \lim_{n\to\infty} z_n = \alpha \cdot 0 = 0.$$

Thus, $\{z_n\}_1^{\infty} + \{w_n\}_1^{\infty}, \alpha\{z_n\}_1^{\infty} \in U$, so U is a subspace of $\mathbb{C}^{\infty}$. □

Solution 1C.3 Let $U = \{f \in \mathbb{R}^{(-4,4)} : f'(-1) = 3f(2)\}$. Since the zero function is differentiable on $(-4, 4)$ and $0'(-1) = 0 = 3(0) = 3 \cdot 0(2)$, $0 \in U$, so U is nonempty. Let $f, g \in U$ and $\alpha \in \mathbb{R}$. Then

$$(f + g)'(-1) = f'(-1) + g'(-1) = 3f(2) + 3g(2) = 3(f(2) + g(2)) = 3(f + g)(2),$$

and

$$(\alpha f)'(-1) = \alpha f'(-1) = \alpha(3f(2)) = 3(\alpha f(2)) = 3(\alpha f)(2).$$

Hence, $f + g, \alpha f \in U$, so U is a subspace of $\mathbb{R}^{(-4,4)}$. □

Solution 1C.4 Define

$$U = \left\{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous and } \int_0^1 f = b\right\}.$$

$\Rightarrow$ Suppose U is a subspace of $\mathbb{R}^{[0,1]}$. Then $0 \in U$, so $\int_0^1 0 = 0 = b$, which implies $b = 0$.
$\Leftarrow$ Suppose $b = 0$. Then

$$U = \left\{f \in \mathbb{R}^{[0,1]} : f \text{ is continuous and } \int_0^1 f = 0\right\}.$$

Since the zero function is continuous on $[0, 1]$ and $\int_0^1 0 = 0$, $0 \in U$, so U is nonempty. Let $f, g \in U$ and $\alpha \in \mathbb{R}$. Then

$$\int_0^1 (f + g) = \int_0^1 f + \int_0^1 g = 0 + 0 = 0,$$

and

$$\int_0^1 (\alpha f) = \alpha \int_0^1 f = \alpha(0) = 0.$$

Therefore, $f + g, \alpha f \in U$, so U is a subspace of $\mathbb{R}^{[0,1]}$. □

Solution 1C.5 No, $\mathbb{R}^2$ is not a subspace of $\mathbb{C}^2$. Let $x = (1,0) \in \mathbb{R}^2$ and $\alpha = i \in \mathbb{C}$. Then $\alpha x = (i,0) \notin \mathbb{R}^2$, so $\mathbb{R}^2$ is not closed under scalar multiplication and is not a subspace of $\mathbb{C}^2$. □

Solution 1C.6 Let U be the subset in question.

(a) Yes, U is a subspace of $\mathbb{R}^3$. To see this, note that $0^3 = 0^3$, so $(0,0,0) \in U$, and U is nonempty. Let $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3) \in U$ and $\alpha \in \mathbb{R}$. Then $u_1^3 = u_2^3$ implies

$$u_1^3 - u_2^3 = (u_1 - u_2)(u_1^2 + u_1 u_2 + u_2^2) = 0.$$

If $u_1 - u_2 = 0$, then $u_1 = u_2$ and $(u_1 + v_1)^3 = (u_2 + v_2)^3$. If $u_1^2 + u_1 u_2 + u_2^2 = 0$, we may complete the square to get

$$u_1^2 + u_1 u_2 + u_2^2 = \left(u_1 + \frac{u_2}{2}\right)^2 + \frac{3u_2^2}{4} = 0.$$

Since a square of a real number is nonnegative, we must have $u_1 + \frac{u_2}{2} = 0$ and $\frac{3u_2^2}{4} = 0$, which implies $u_1 = u_2 = 0$. In either case, $u_1^3 = u_2^3$ implies $u_1 = u_2$. A similar argument shows that $v_1^3 = v_2^3$ implies $v_1 = v_2$. Hence,

$$(u_1 + v_1)^3 = (u_2 + v_2)^3,$$

and

$$(\alpha u_1)^3 = \alpha^3 u_1^3 = \alpha^3 u_2^3 = (\alpha u_2)^3.$$

Therefore, $u + v \in U$ and $\alpha u \in U$, so U is a subspace of $\mathbb{R}^3$.

(b) No, U is not a subspace of $\mathbb{C}^3$. To see this, suppose $z^3 = 1$ for some $z \in \mathbb{C}$. Then $z^3 - 1 = (z-1)(z^2 + z + 1) = 0$. Then $z = 1$ or $z^2 + z + 1 = 0$. If $z^2 + z + 1 = 0$, we may use the quadratic formula to obtain

$$z = \frac{-1 \pm \sqrt{1-4}}{2} = \frac{-1 \pm i\sqrt{3}}{2}.$$

Let

$$\omega_1 = \frac{-1 + i\sqrt{3}}{2} \quad \text{and} \quad \omega_2 = \frac{-1 - i\sqrt{3}}{2}.$$

Then $\omega_1^3 = \omega_2^3 = 1$, and $\omega_1 + \omega_2 = -1$. Let $u = (\omega_1, 1, 0), v = (\omega_2, 1, 0) \in U$. Then $u + v = (\omega_1 + \omega_2, 1 + 1, 0 + 0) = (-1, 2, 0)$. Since $(-1)^3 = -1 \neq 8 = 2^3$, $u + v \notin U$. Hence, U is not closed under addition and is not a subspace of $\mathbb{C}^3$. □

Solution 1C.7 Consider the subset $\mathbb{Q}^2$ of $\mathbb{R}^2$. Then $\mathbb{Q}^2$ is nonempty because $(0,0) \in \mathbb{Q}^2$, closed under addition since the sum of any two rational numbers is rational, and closed under taking additive inverses since the additive inverse of a rational number is rational. However, for any irrational number $r \in \mathbb{R}$, the vector $(r,0) = r(1,0) \in \mathbb{R}^2$ is not in $\mathbb{Q}^2$. Hence, $\mathbb{Q}^2$ is not closed under scalar multiplication and is not a subspace of $\mathbb{R}^2$. Therefore, the statement is false. □

Solution 1C.8 Consider the subset

$$U = \{(x, 0) : x \in \mathbb{R}\} \cup \{(0, y) : y \in \mathbb{R}\}.$$

Then U is nonempty since $(1, 0) \in U$, and it is closed under scalar multiplication since any scalar multiple of $(1, 0)$ or $(0, 1)$ is still in U. However, U is not closed under addition since $(1, 0) + (0, 1) = (1, 1) \notin U$. Hence, U is not a subspace of $\mathbb{R}^2$. □

Solution 1C.9 It is not a subspace of $\mathbb{R}^{\mathbb{R}}$. Consider the periodic functions $f(x) = \cos x$ and $g(x) = \cos \sqrt{2}x$. Then the sum $f + g$ attains a maximum value of 2 at 0. If $f + g$ were periodic with period p, then $(f + g)(p) = (f + g)(0) = 2$. This implies $\cos p = \cos p\sqrt{2} = 1$. Then $p = 2\pi m$ and $p\sqrt{2} = 2\pi n$ for positive integers m and n. Since $p\sqrt{2} = 2\pi n$ implies $p = \sqrt{2}\pi n$, we have

$$2\pi m = \sqrt{2}\pi n \implies \sqrt{2} = \frac{n}{m},$$

which is a contradiction, since $\sqrt{2}$ is irrational. Then $f + g$ is not periodic. □

Solution 1C.10 Let $U = V_1 \cap V_2$. Since both V_1 and V_2 are subspaces of V, they both contain the zero vector of V. Hence, $0 \in U$. Let $u, v \in U$. Then $u, v \in V_1$ and $u, v \in V_2$. Since both V_1 and V_2 are subspaces, they are closed under addition. Therefore, $u + v \in V_1$ and $u + v \in V_2$, which implies that $u + v \in U$. Let $a \in \mathbb{F}$. Then since both V_1 and V_2 are closed under scalar multiplication, we have that $au \in V_1$ and $au \in V_2$, which implies that $au \in U$. Therefore, U is a subspace of V. □

Solution 1C.11 Let $\mathscr{C}$ be a collection of subspaces of V, and let

$$U = \bigcap_{W \in \mathscr{C}} W.$$

Since each subspace in $\mathscr{C}$ contains the zero vector, $0 \in U$. Let $u, v \in U$. Then $u, v \in W$ for all $W \in \mathscr{C}$. Since each W is a subspace, they are closed under addition, so $u+v \in W$ for all $W \in \mathscr{C}$, which implies $u + v \in U$. Let $a \in \mathbb{F}$. Then $au \in W$ for all $W \in \mathscr{C}$, which implies $au \in U$. Therefore, U is a subspace of V. □

Solution 1C.12 Let U, W be subspaces of V. Suppose $U \subseteq W$. Then $U \cup W = W$, which is a subspace of V. Similarly, if $W \subseteq U$, then $U \cup W = U$, which is a subspace of V.

Conversely, suppose $U \cup W$ is a subspace of V. If neither $U \subseteq W$ nor $W \subseteq U$, then there exist $u \in U \setminus W$ and $w \in W \setminus U$. Since $U \cup W$ is a subspace, it must be closed under addition. Then $u + w \in U \cup W$. However, $u + w \notin U$ because if it were, then $w = (u + w) - u \in U$, which is a contradiction. Similarly, $u + w \notin W$ because if it were, then $u = (u+w)-w \in W$, which is also a contradiction. Hence, $u+w \notin U \cup W$, which contradicts the assumption that $U \cup W$ is a subspace. Therefore, one of the subspaces must be contained in the other. □

Solution 1C.13

Appendix. *An alternate proof is given at the end of this solution.*

$\Rightarrow$ Let U_1, U_2, U_3 be subspaces of V. Without loss of generality, suppose $U_2 \subseteq U_1$ and $U_3 \subseteq U_1$. Then $U_1 \cup U_2 \cup U_3 = U_1$, which is a subspace of V.

$\Leftarrow$ Conversely, suppose $U_1 \cup U_2 \cup U_3$ is a subspace of V. Suppose that some pair satisfies containment, say $U_3 \subseteq U_2$. Let $W = U_2 \cup U_3 = U_2$. Then $U_1 \cup W = U_1 \cup U_2 \cup U_3$. By Exercise 1C12, either $U_1 \subseteq U_2$ or $U_2 \subseteq U_1$. If the former occurs, then $U_1 \cup U_2 = U_2$, which contains U_3. If the latter occurs, then $U_1 \cup U_2 = U_1$. Since $U_3 \subseteq U_2$, then $U_3 \subseteq U_1$. Therefore, one of the subspaces must contain the other two.

Now suppose no subspace is contained in another. Then there exist $u \in U_2 \setminus U_3$ and $v \in U_3 \setminus U_2$. Since $U_1 \cup U_2 \cup U_3$ is a subspace, $u + \alpha v \in U_1 \cup U_2 \cup U_3$ for any nonzero $\alpha \in \mathbb{F}$. If $u + \alpha v \in U_2$, then $\alpha v = (u + \alpha v) - u \in U_2$, which implies $v \in U_2$, a contradiction. If $u + \alpha v \in U_3$, then $u = (u + \alpha v) - \alpha v \in U_3$, a contradiction. Then $u + \alpha v \in U_1$ for all nonzero $\alpha \in \mathbb{F}$. Now choose nonzero $\beta, \gamma \in \mathbb{F}$ with $\beta \neq \gamma$. Then $u + \beta v, u + \gamma v \in U_1$, so

$$(u + \beta v) - (u + \gamma v) = (\beta - \gamma)v \in U_1.$$

Since $\beta - \gamma \neq 0$, we have $v \in U_1$. Similarly, $(u + \beta v) - \beta v = u \in U_1$. Since $u \in U_2 \setminus U_3$ and $v \in U_3 \setminus U_2$ were arbitrary, $U_2 \setminus U_3 \subseteq U_1$ and $U_3 \setminus U_2 \subseteq U_1$.

Now let $w \in U_2 \cap U_3$. Since $U_2 \not\subseteq U_3$, there exists $x \in U_2 \setminus U_3 \subseteq U_1$. Then $w + x \in U_2$. If $w + x \in U_3$, then $x = (w + x) - w \in U_3$, a contradiction. Hence, $w + x \in U_2 \setminus U_3 \subseteq U_1$, so $w = (w + x) - x \in U_1$. Therefore, $U_2 \cap U_3 \subseteq U_1$, which gives $U_2 = (U_2 \setminus U_3) \cup (U_2 \cap U_3) \subseteq U_1$ and similarly $U_3 \subseteq U_1$, contradicting our assumption. Hence, one of the subspaces must contain the other two.

Alternate proof. Coolempire93. Conversely, suppose that $U_1 \cup U_2 \cup U_3$ is a subspace of V but none of the subspaces contains the other two. Let $W_1 = U_2 \cup U_3$ and $W_2 = U_1 \cup U_3$. By the assumption, we cannot have $W_1 \subseteq U_1$ or $W_2 \subseteq U_2$, so there exist $x \in U_1 \setminus W_1$ and $y \in U_2 \setminus W_2$. Now, since $U_1 \cup U_2 \cup U_3$ is a vector space, we have $x + \alpha y \in U_1 \cup U_2 \cup U_3$ for any nonzero $\alpha \in \mathbb{F}$. But if $x + \alpha y \in U_1$, then $(x + \alpha y) - x = \alpha y \in U_1$, and as α is nonzero, this means that $y \in U_1$, a contradiction. Similarly, if $x + \alpha y \in U_2$, then $(x + \alpha y) - \alpha y = x \in U_2$, a contradiction. Therefore, we must have $x + \alpha y \in U_3$ for all nonzero α to satisfy the closure of the vector space.

As such, $x + y \in U_3$ and $x - y \in U_3$; that is, $(x + y) + (x - y) = 2x \in U_3$ and $(x + y) - (x - y) = 2y \in U_3$, contradicting the original definitions of x and y[†]. Therefore, one of the subspaces must contain the other two.

[†] *Final remark.* This is why the field must have more than two elements. Since part of the field axioms is that $0 \neq 1$, if the field only has two elements, it must be that $1 = -1$ (consider addition mod 2 as the addition operation), so $2x \in U_3$ would not mean that $x \in U_3$, as $2x = 0x = 0$. □

Solution 1C.14 Let $u = (x, -x, 2x) \in U$ and $w = (y, y, 2y) \in W$. Then

$$u + w = (x + y, -x + y, 2x + 2y) = (x + y, y - x, 2(x + y)).$$

Observe that $x + y$ can take any value in $\mathbb{F}$, and $y - x$ can also take any value in $\mathbb{F}$. Hence, we can write

$$U + W = \{(a, b, 2a) \in \mathbb{F}^3 : a, b \in \mathbb{F}\}.$$

In words, $U + W$ is the set of all vectors in $\mathbb{F}^3$ whose third component is twice the first component, and whose second component can be any value in $\mathbb{F}$. □

Solution 1C.15 Let $u_1, u_2 \in U$. Then $u_1 + u_2 \in U$ since U is closed under addition. Hence, $U + U \subseteq U$. Conversely, let $u \in U$. Then $u = u + 0$, where 0 is the additive identity in U. Since $0 \in U$, we have that $u \in U + U$. Hence, $U \subseteq U + U$. Therefore, $U + U = U$. □

Solution 1C.16 Yes, the operation of addition on subspaces of V is commutative. Let $u \in U$ and $w \in W$. Then $u + w \in U + W$. Because addition in V is commutative, we have that $u + w = w + u$, where $w \in W$ and $u \in U$. Hence, $w + u \in W + U$. Therefore, every element of $U + W$ is also an element of $W + U$, which implies that $U + W \subseteq W + U$. By a similar argument, we can show that $W + U \subseteq U + W$. Hence, $U + W = W + U$. □

Solution 1C.17 Yes, the operation of addition on subspaces of V is associative. Let $u_1 \in V_1$, $u_2 \in V_2$, and $u_3 \in V_3$. Then $(u_1 + u_2) + u_3 \in (V_1 + V_2) + V_3$. Because addition in V is associative, we have that $(u_1 + u_2) + u_3 = u_1 + (u_2 + u_3)$, where $u_1 \in V_1$, $u_2 \in V_2$, and $u_3 \in V_3$. Hence, $u_1 + (u_2 + u_3) \in V_1 + (V_2 + V_3)$. Therefore, every element of $(V_1 + V_2) + V_3$ is also an element of $V_1 + (V_2 + V_3)$, which implies that $(V_1 + V_2) + V_3 \subseteq V_1 + (V_2 + V_3)$. By a similar argument, we can show that $V_1 + (V_2 + V_3) \subseteq (V_1 + V_2) + V_3$. Hence, $(V_1 + V_2) + V_3 = V_1 + (V_2 + V_3)$. □

Solution 1C.18 Yes, the operation of addition on the subspaces of V has an additive identity. The additive identity is the zero subspace $\{0\}$, which contains only the zero vector of V. For any subspace U of V, we have that $U + \{0\} = U$.
If a subspace U of V had an additive inverse W, then we would have that $U + W = \{0\}$. However, $U \subseteq U + W$ since $u = u + 0$ for any $u \in U$. Hence, $U + W = \{0\}$ implies that $U = \{0\}$. Therefore, the only subspace of V that has an additive inverse is the zero subspace $\{0\}$. □

Solution 1C.19 The statement is false. Consider the vector space $\mathbb{R}^2$ and the subspaces

$$V_1 = \{(x, 0) \in \mathbb{R}^2 : x \in \mathbb{R}\}, \quad V_2 = \{(0, y) \in \mathbb{R}^2 : y \in \mathbb{R}\}, \quad U = \{(z, z) \in \mathbb{R}^2 : z \in \mathbb{R}\}.$$

Then $V_1 + U = \mathbb{R}^2$ and $V_2 + U = \mathbb{R}^2$, but $V_1 \neq V_2$. □

Solution 1C.20 Let

$$W = \{(0, w_2, 0, w_4) \in \mathbb{F}^4 : w_2, w_4 \in \mathbb{F}\}.$$

Suppose $(x_1, x_2, x_3, x_4) \in U \cap W$. Since $(x_1, x_2, x_3, x_4) \in U$, we have that $x_1 = x_2$ and $x_3 = x_4$. Moreover, since $(x_1, x_2, x_3, x_4) \in W$, we have that $x_1 = 0$ and $x_3 = 0$. Hence, $x_1 = x_2 = x_3 = x_4 = 0$, so $U \cap W = \{0\}$. Let $(y_1, y_2, y_3, y_4) \in \mathbb{F}^4$. Then we can write

$$(y_1, y_2, y_3, y_4) = (y_1, y_1, y_3, y_3) + (0, y_2 - y_1, 0, y_4 - y_3),$$

where $(y_1, y_1, y_3, y_3) \in U$ and $(0, y_2 - y_1, 0, y_4 - y_3) \in W$. Hence, every $x \in \mathbb{F}^4$ can be uniquely written as $x = u + w$ for some $u \in U$ and $w \in W$, so $\mathbb{F}^4 = U \oplus W$. □

Solution 1C.21 Let

$$W = \{(0, 0, w_3, w_4, w_5) \in \mathbb{F}^5 : w_3, w_4, w_5 \in \mathbb{F}\}.$$

Suppose $(x_1, x_2, x_3, x_4, x_5) \in U \cap W$. Since $(x_1, x_2, x_3, x_4, x_5) \in U$, we have that $x_3 = x_1 + x_2$, $x_4 = x_1 - x_2$, and $x_5 = 2x_1$. Moreover, since $(x_1, x_2, x_3, x_4, x_5) \in W$, we have that $x_1 = 0$ and $x_2 = 0$. Hence, $x_1 = x_2 = x_3 = x_4 = x_5 = 0$, so $U \cap W = \{0\}$. Let $(y_1, y_2, y_3, y_4, y_5) \in \mathbb{F}^5$. Then we can write

$$(y_1, y_2, y_3, y_4, y_5) = (y_1, y_2, y_1+y_2, y_1-y_2, 2y_1)+(0, 0, y_3-(y_1+y_2), y_4-(y_1-y_2), y_5-2y_1),$$

where $(y_1, y_2, y_1 + y_2, y_1 - y_2, 2y_1) \in U$ and $(0, 0, y_3 - (y_1 + y_2), y_4 - (y_1 - y_2), y_5 - 2y_1) \in W$. Hence, every $x \in \mathbb{F}^5$ can be written as $x = u + w$ for some $u \in U$ and $w \in W$, so $\mathbb{F}^5 = U \oplus W$. □

Solution 1C.22 Let

$$W_1 = \{(0, 0, w_3, 0, 0) \in \mathbb{F}^5 : w_3 \in \mathbb{F}\},$$
$$W_2 = \{(0, 0, 0, w_4, 0) \in \mathbb{F}^5 : w_4 \in \mathbb{F}\},$$
$$W_3 = \{(0, 0, 0, 0, w_5) \in \mathbb{F}^5 : w_5 \in \mathbb{F}\}.$$

Consider

$$W = \{(0, 0, z_3, z_4, z_5) \in \mathbb{F}^5 : z_3, z_4, z_5 \in \mathbb{F}\}.$$

It is straightforward to show that $W = W_1 \oplus W_2 \oplus W_3$. By Exercise 1C21, we have that $\mathbb{F}^5 = U \oplus W$. By uniqueness of the decomposition of $v \in \mathbb{F}^5$ as $v = u + w$ for some $u \in U$ and $w \in W$, and uniqueness of the decomposition of $w \in W$ as $w = w_1 + w_2 + w_3$ for some $w_1 \in W_1$, $w_2 \in W_2$, and $w_3 \in W_3$, we have that $\mathbb{F}^5 = U \oplus W_1 \oplus W_2 \oplus W_3$. □

Solution 1C.23 See Exercise 1C19 for the definitions of V_1, V_2, U, and let $V = \mathbb{R}^2$. Suppose $(x_1, y_1) \in V_1 \cap U$. Since $(x_1, y_1) \in U$, we have that $x_1 = y_1$. Since $(x_1, y_1) \in V_1$, we have that $y_1 = 0$. Hence, $x_1 = y_1 = 0$, so $V_1 \cap U = \{0\}$. Similarly, if $(x_2, y_2) \in V_2 \cap U$, then $x_2 = y_2 = 0$, so $V_2 \cap U = \{0\}$. Let $(x, y) \in V$. Then we can write

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

where $(x - y, 0) \in V_1$ and $(y, y) \in U$. Additionally, we can write

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

where $(0, y - x) \in V_2$ and $(x, x) \in U$. Hence, $V = V_1 \oplus U$ and $V = V_2 \oplus U$, but $V_1 \neq V_2$. □

Solution 1C.24 Let $f(x) = 0$ for all $x \in \mathbb{R}$. Since $f(-x) = f(x) = 0$, and $f(-x) = -f(x) = 0$, we have $f(x) \in V_e$ and $f(x) \in V_o$. Now suppose $f_1, f_2 \in V_e$ and $\alpha \in \mathbb{R}$. Then

$$(f_1 + f_2)(-x) = f_1(-x) + f_2(-x) = f_1(x) + f_2(x) = (f_1 + f_2)(x)$$

Hence, $f_1 + f_2 \in V_e$. Also,

$$(\alpha f_1)(-x) = \alpha f_1(-x) = \alpha f_1(x) = (\alpha f_1)(x),$$

so $\alpha f_1 \in V_e$. Therefore, V_e is a subspace of $\mathbb{R}^{\mathbb{R}}$.
Similarly, if $g_1, g_2 \in V_o$ and $\beta \in \mathbb{R}$, then

$$(g_1 + g_2)(-x) = g_1(-x) + g_2(-x) = -g_1(x) - g_2(x) = -(g_1 + g_2)(x)$$

Hence, $g_1 + g_2 \in V_o$. Also,

$$(\beta g_1)(-x) = \beta g_1(-x) = \beta(-g_1(x)) = -(\beta g_1)(x),$$

so $\beta g_1 \in V_o$. Therefore, V_o is a subspace of $\mathbb{R}^{\mathbb{R}}$.

We now show $\mathbb{R}^{\mathbb{R}} = V_e + V_o$. Let $g \in \mathbb{R}^{\mathbb{R}}$. Define the functions g_e and g_o by

$$g_e(x) = \frac{g(x) + g(-x)}{2}, \quad g_o(x) = \frac{g(x) - g(-x)}{2}.$$

Then

$$g_e(-x) = \frac{g(-x) + g(x)}{2} = \frac{g(x) + g(-x)}{2} = g_e(x),$$

and

$$g_o(-x) = \frac{g(-x) - g(x)}{2} = -\frac{g(x) - g(-x)}{2} = -g_o(x).$$

Hence, $g_e \in V_e$ and $g_o \in V_o$. Moreover,

$$g(x) = g_e(x) + g_o(x),$$

so $g \in V_e + V_o$.

Finally, if $h \in V_e \cap V_o$, then $h(-x) = h(x)$ and $h(-x) = -h(x)$ for all $x \in \mathbb{R}$. Hence, $h(x) = -h(x)$, so $h(x) = 0$ for all $x \in \mathbb{R}$. Therefore, $V_e \cap V_o = \{0\}$, and we conclude that $\mathbb{R}^{\mathbb{R}} = V_e \oplus V_o$. □

Chapter 2

§2A

Solution 2A.1 Let $U = \{(x, y, z) \in \mathbb{F}^3 : x + y + z = 0\}$. Rewriting $z = -x - y$, we have

$$U = \{(x, y, -x - y) : x, y \in \mathbb{F}\}.$$

Consider the list of vectors

$$v_1 = (1, 0, -1), \quad v_2 = (0, 1, -1), \quad v_3 = (2, -1, -1), \quad v_4 = (1, 1, -2).$$

For any $(z, w, -z - w) \in U$, we may write it as $zv_1 + wv_2$. Therefore, v_1, v_2, v_3, v_4 span U. □

Solution 2A.2 Let $U = \{v_1, v_2, v_3, v_4\}$ and $W = \{v_1 - v_2, v_2 - v_3, v_3 - v_4, v_4\}$. Clearly, $v_4 \in \operatorname{span} U$. Then we may write

$$\begin{aligned} v_3 &= (v_3 - v_4) + v_4 \in \operatorname{span} W, \\ v_2 &= (v_2 - v_3) + v_3 \in \operatorname{span} W, \\ v_1 &= (v_1 - v_2) + v_2 \in \operatorname{span} W. \end{aligned}$$

Thus, $v_1, v_2, v_3, v_4 \in \operatorname{span} W$, and so $V = \operatorname{span} U \subseteq \operatorname{span} W$. Since W is a list of vectors in V, we have $\operatorname{span} W \subseteq V$. Therefore, $\operatorname{span} W = V$, and so W spans V. □

Solution 2A.3 This is a generalization of Exercise 2A2 and a direct consequence of the fact that $v_k = w_k - w_{k-1}$ for $k \in \{2, \ldots, m\}$ and $v_1 = w_1$. Since each v_k is a linear combination of $w_1, \ldots, w_k$, we have $\operatorname{span}(v_1, \ldots, v_m) \subseteq \operatorname{span}(w_1, \ldots, w_m)$. On the other hand, each w_k is a linear combination of $v_1, \ldots, v_k$, so $\operatorname{span}(w_1, \ldots, w_m) \subseteq \operatorname{span}(v_1, \ldots, v_m)$. Therefore, $\operatorname{span}(v_1, \ldots, v_m) = \operatorname{span}(w_1, \ldots, w_m)$. □

Solution 2A.4

(a) Let v be the vector in the list. If $v = 0$, then the list is linearly dependent, since $0 \cdot v = 0$. If $v \neq 0$, then the only way to write 0 as a linear combination of v is to take the scalar multiple to be 0. Thus, the list is linearly independent.

(b) Let v_1 and v_2 be the vectors in the list. If v_1 is a scalar multiple of v_2, then there exists a scalar λ such that $v_1 = \lambda v_2$. Then we may write

$$0 = v_1 - \lambda v_2,$$

which shows that the list is linearly dependent. Conversely, if the list is linearly dependent, then there exist scalars α and β, not both zero, such that

$$\alpha v_1 + \beta v_2 = 0.$$

If $\alpha \neq 0$, then

$$v_1 = -\frac{\beta}{\alpha} v_2.$$

Similarly, if $\beta \neq 0$, then

$$v_2 = -\frac{\alpha}{\beta} v_1.$$

Thus, the list is linearly independent if and only if neither vector is a scalar multiple of the other.

□

Solution 2A.5 To say that the list is not linearly independent is to say that there exist scalars α, β, γ, not all zero, such that

$$\alpha(3,1,4) + \beta(2,-3,5) + \gamma(5,9,t) = (0,0,0).$$

This is equivalent to the following system of equations:

$$\begin{aligned} 0 &= 3\alpha + 2\beta + 5\gamma, \\ 0 &= \alpha - 3\beta + 9\gamma, \\ 0 &= 4\alpha + 5\beta + t\gamma. \end{aligned}$$

From the second equation, we obtain $\alpha = 3\beta - 9\gamma$. Substituting this into the first equation, we have

$$\begin{aligned} 0 &= 3(3\beta - 9\gamma) + 2\beta + 5\gamma \\ &= 11\beta - 22\gamma. \end{aligned}$$

Thus, $\beta = 2\gamma$. Substituting this into the expression for α, we have $\alpha = 3(2\gamma) - 9\gamma = -3\gamma$. Substituting these values into the third equation, we have

$$\begin{aligned} 0 &= 4(-3\gamma) + 5(2\gamma) + t\gamma \\ &= -12\gamma + 10\gamma + t\gamma \\ &= (t-2)\gamma. \end{aligned}$$

Since $\gamma \neq 0$, we must have $t - 2 = 0$, or $t = 2$. Then the list is not linearly independent when $t = 2$. □

Solution 2A.6 We may proceed similarly to Exercise 2A5 to obtain the following system of equations:

$$\begin{aligned} 0 &= 2\alpha + \beta + 7\gamma, \\ 0 &= 3\alpha - \beta + 3\gamma, \\ 0 &= \alpha + 2\beta + c\gamma. \end{aligned}$$

The second equation gives us $\beta = 3\alpha + 3\gamma$. Substituting this into the first equation, we have

$$\begin{aligned} 0 &= 2\alpha + (3\alpha + 3\gamma) + 7\gamma \\ &= 5\alpha + 10\gamma. \end{aligned}$$

Thus, $\alpha = -2\gamma$. Substituting this into the expression for β, we have $\beta = 3(-2\gamma) + 3\gamma = -3\gamma$. Substituting these values into the third equation, we have

$$\begin{aligned} 0 &= (-2\gamma) + 2(-3\gamma) + c\gamma \\ &= -2\gamma - 6\gamma + c\gamma \\ &= (c-8)\gamma. \end{aligned}$$

Since $\gamma \neq 0$, we must have $c - 8 = 0$, or $c = 8$. Then the list is linearly dependent when $c = 8$.

Conversely, if $c \neq 8$, then $(c-8)\gamma = 0$ implies $\gamma = 0$. Then $\alpha = -2\gamma = 0$ and $\beta = -3\gamma = 0$, so the only solution to the system of equations is $\alpha = \beta = \gamma = 0$. Thus, the list is linearly independent when $c \neq 8$. □

Solution 2A.7

(a) Suppose $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha(1+i) + \beta(1-i) = 0.$$

Then we have

$$(\alpha + \beta) + (\alpha - \beta)i = 0.$$

Since $0 = 0 + 0i$, we must have $\alpha + \beta = 0$ and $\alpha - \beta = 0$. Solving this system of equations, we find that $\alpha = \beta = 0$. Thus, the list is linearly independent.

(b) Suppose $a + bi, c + di \in \mathbb{C}$ such that

$$(a+bi)(1+i) + (c+di)(1-i) = 0.$$

Then we have

$$(a - b + c + d) + (a + b - c + d)i = 0.$$

Then we have the system of equations

$$0 = a - b + c + d,$$
$$0 = a + b - c + d.$$

Adding the two equations results in $0 = 2a + 2d$, or $a = -d$. Substituting this into the first equation, we obtain $b = c$. Then we have infinitely many nontrivial solutions to the system of equations, such as $a = 1, b = 0, c = 0, d = -1$. Thus, the list is linearly dependent.

□

Solution 2A.8 Let $\alpha, \beta, \gamma, \delta \in \mathbb{F}$ such that

$$\alpha(v_1 - v_2) + \beta(v_2 - v_3) + \gamma(v_3 - v_4) + \delta v_4 = 0.$$

Then we have

$$\alpha v_1 + (-\alpha + \beta)v_2 + (-\beta + \gamma)v_3 + (-\gamma + \delta)v_4 = 0.$$

Since v_1, v_2, v_3, v_4 is linearly independent, we must have

$$0 = \alpha,$$
$$0 = -\alpha + \beta,$$
$$0 = -\beta + \gamma,$$
$$0 = -\gamma + \delta.$$

From the first equation, we have $\alpha = 0$. Substituting this into the second equation gives us $\beta = 0$. Substituting this into the third equation gives us $\gamma = 0$. Finally, substituting this into the fourth equation gives us $\delta = 0$. Thus, the only solution to the original equation is $\alpha = \beta = \gamma = \delta = 0$, and so the list is linearly independent. □

Solution 2A.9 Let $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that

$$\alpha_1(5v_1 - 4v_2) + \alpha_2 v_2 + \cdots + \alpha_m v_m = 0.$$

Then we have

$$5\alpha_1 v_1 + (-4\alpha_1 + \alpha_2)v_2 + \alpha_3 v_3 + \cdots + \alpha_m v_m = 0.$$

Since $v_1, v_2, \ldots, v_m$ is linearly independent, we must have

$$\begin{aligned} 0 &= 5\alpha_1, \\ 0 &= -4\alpha_1 + \alpha_2, \\ 0 &= \alpha_3, \\ &\vdots \\ 0 &= \alpha_m. \end{aligned}$$

From the first equation, we have $\alpha_1 = 0$. Substituting this into the second equation gives us $\alpha_2 = 0$. The remaining equations give us $\alpha_k = 0$ for $k = 3, \ldots, m$. Thus, the only solution to the original equation is $\alpha_k = 0$ for $k = 1, \ldots, m$, and so the list is linearly independent. □

Solution 2A.10 Let $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that

$$\alpha_1(\lambda v_1) + \cdots + \alpha_m(\lambda v_m) = 0.$$

Then we have

$$\lambda(\alpha_1 v_1 + \cdots + \alpha_m v_m) = 0.$$

Since $\lambda \neq 0$, we must have

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0.$$

Since $v_1, v_2, \ldots, v_m$ is linearly independent, we must have $\alpha_k = 0$ for $k = 1, \ldots, m$. Thus, the only solution to the original equation is $\alpha_k = 0$ for $k = 1, \ldots, m$, and so the list is linearly independent. □

Solution 2A.11 Consider the following counterexample in $\mathbb{R}^2$:

$$v_1 = (1, 0), \quad v_2 = (0, 1), \quad w_1 = (0, 1), \quad w_2 = (1, 0).$$

Then v_1, v_2 and w_1, w_2 are linearly independent lists of vectors in $\mathbb{R}^2$. However,

$$v_1 + w_1 = (1, 1), \quad v_2 + w_2 = (1, 1),$$

and so the list $v_1 + w_1, v_2 + w_2$ is linearly dependent. Thus, the statement is false. □

Solution 2A.12 If $w = 0$, then $w \in \text{span}(v_1, \ldots, v_m)$ and we are done. Otherwise, assume $w \neq 0$, and let $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$, not all zero, such that

$$\alpha_1(v_1 + w) + \cdots + \alpha_m(v_m + w) = 0.$$

Then we have

$$(\alpha_1 v_1 + \cdots + \alpha_m v_m) + (\alpha_1 + \cdots + \alpha_m)w = 0.$$

If $\alpha_1 + \cdots + \alpha_m = 0$, then we have

$$\alpha_1 v_1 + \cdots + \alpha_m v_m = 0,$$

which contradicts the linear independence of $v_1, \ldots, v_m$. Thus, we must have $\alpha_1 + \cdots + \alpha_m \neq 0$. Let $\beta = \alpha_1 + \cdots + \alpha_m$. Then we may write

$$w = -\frac{\alpha_1}{\beta} v_1 - \cdots - \frac{\alpha_m}{\beta} v_m.$$

Thus, w is a linear combination of $v_1, \ldots, v_m$, and so $w \in \text{span}(v_1, \ldots, v_m)$. □

Solution 2A.13 $\Rightarrow$ We proceed by contrapositive. Suppose $w \in \text{span}(v_1, \ldots, v_m)$, but $v_1, \ldots, v_m, w$ is linearly independent. Then w is a linear combination of $v_1, \ldots, v_m$, contradicting the linear independence of $v_1, \ldots, v_m, w$. Hence, $w \notin \text{span}(v_1, \ldots, v_m)$. $\Leftarrow$ Conversely, suppose $w \notin \text{span}(v_1, \ldots, v_m)$, but $v_1, \ldots, v_m, w$ is linearly dependent. By the linear dependence lemma, there exists $k \in \{1, \ldots, m\}$ such that $v_k \in \text{span}(v_1, \ldots, v_{k-1})$. This contradicts the linear independence of $v_1, \ldots, v_m$. Hence, $v_1, \ldots, v_m, w$ must be linearly independent. □

Solution 2A.14 Suppose $v_1, \ldots, v_m$ is linearly independent. Let $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that

$$\alpha_1 w_1 + \cdots + \alpha_m w_m = 0.$$

Substituting $w_k = v_1 + \cdots + v_k$ for each k, we get

$$\alpha_1 v_1 + \alpha_2(v_1 + v_2) + \alpha_3(v_1 + v_2 + v_3) + \cdots + \alpha_m(v_1 + \cdots + v_m) = 0.$$

Then we may rewrite this as

$$(\alpha_1 + \cdots + \alpha_m)v_1 + (\alpha_2 + \cdots + \alpha_m)v_2 + \cdots + (\alpha_{m-1} + \alpha_m)v_{m-1} + \alpha_m v_m = 0.$$

Since $v_1, \ldots, v_m$ is linearly independent, we must have $\alpha_1 = \cdots = \alpha_m = 0$. Hence, $w_1, \ldots, w_m$ is linearly independent.
Conversely, suppose $w_1, \ldots, w_m$ is linearly independent. Let $\beta_1, \ldots, \beta_m \in \mathbb{F}$ such that

$$\beta_1 v_1 + \cdots + \beta_m v_m = 0.$$

Noting that $v_1 = w_1$ and $v_k = w_k - w_{k-1}$ for $k \in \{2, \ldots, m\}$, we may rewrite the equation as

$$\begin{aligned} \beta_1 w_1 + \beta_2(w_2 - w_1) + \cdots + \beta_m(w_m - w_{m-1}) &= 0, \\ (\beta_1 - \beta_2)w_1 + (\beta_2 - \beta_3)w_2 + \cdots + (\beta_{m-1} - \beta_m)w_{m-1} + \beta_m w_m &= 0. \end{aligned}$$

Since $w_1, \ldots, w_m$ is linearly independent, we find that the system of equations

$$\begin{aligned} 0 &= \beta_1 - \beta_2, \\ 0 &= \beta_2 - \beta_3, \\ &\vdots \\ 0 &= \beta_{m-1} - \beta_m \\ 0 &= \beta_m, \end{aligned}$$

has the unique solution $\beta_1 = \cdots = \beta_m = 0$. Hence, $v_1, \ldots, v_m$ is linearly independent. □

Solution 2A.15 Since $\mathbb{F}[X]_{\leq 4}$ can be spanned by the list $1, x, x^2, x^3, x^4$, and the length of a linearly independent list cannot exceed the length of a spanning list, there does not exist a list of six polynomials that is linearly independent in $\mathbb{F}[X]_{\leq 4}$. □

Solution 2A.16 Since the list $1, x, x^2, x^3, x^4$ is a linearly independent list in $\mathbb{F}[X]_{\leq 4}$ and has length 5, and the length of a spanning list cannot be less than the length of a linearly independent list, no list of four polynomials can span $\mathbb{F}[X]_{\leq 4}$. □

Solution 2A.17 Suppose V is infinite-dimensional. We proceed by induction to construct a sequence of vectors $v_1, v_2, \ldots$ in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m. Since V is infinite-dimensional, there exists a nonzero vector $v_1 \in V$. Suppose we have constructed a linearly independent list of vectors $v_1, \ldots, v_m$ in V. If every $v \in V$ is a linear combination of $v_1, \ldots, v_m$, then $V = \operatorname{span}(v_1, \ldots, v_m)$, which contradicts the assumption that V is infinite-dimensional. Thus, there exists a vector $v_{m+1} \in V$ that is not in $\operatorname{span}(v_1, \ldots, v_m)$. Then by Exercise 2A13, the list $v_1, \ldots, v_m, v_{m+1}$ is linearly independent. By induction, we can construct a sequence of vectors $v_1, v_2, \ldots$ in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m.

Conversely, suppose there exists a sequence of vectors $v_1, v_2, \ldots$ in V such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m. Then for every positive integer m, there exists a linearly independent list of vectors in V of length m. Since there is no maximum length for a linearly independent list of vectors in V, no list can span V. Therefore, V is infinite-dimensional. □

Solution 2A.18 By Exercise 2A17, it suffices to construct a sequence of vectors $v_1, v_2, \ldots$ in $\mathbb{F}^\infty$ such that $v_1, \ldots, v_m$ is linearly independent for every positive integer m. Let v_k be the vector in $\mathbb{F}^\infty$ whose k-th coordinate is 1 and all other coordinates are 0. Explicitly, we have

$$v_1 = (1, 0, 0, \ldots), \quad v_2 = (0, 1, 0, \ldots), \quad v_3 = (0, 0, 1, \ldots), \quad \ldots$$

Then for any positive integer m, the list $v_1, \ldots, v_m$ is linearly independent, since the only way to write the zero vector as a linear combination of $v_1, \ldots, v_m$ is to take all coefficients to be zero. Thus, by Exercise 2A17, $\mathbb{F}^\infty$ is infinite-dimensional. □

Solution 2A.19 By Exercise 2A17, it suffices to construct a sequence of continuous real-valued functions $f_1, f_2, \ldots$ on the interval $[0, 1]$ such that $f_1, \ldots, f_m$ is linearly independent for every positive integer m. Let $f_k(x) = x^{k-1}$ for each positive integer k. Then for any positive integer m, let $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that

$$\alpha_1 f_1(x) + \cdots + \alpha_m f_m(x) = 0$$

for all $x \in [0, 1]$. This is equivalent to the polynomial equation

$$\alpha_1 + \alpha_2 x + \cdots + \alpha_m x^{m-1} = 0$$

for all $x \in [0, 1]$. Since a polynomial of degree at most $m - 1$ can have at most $m - 1$ roots unless it is the zero polynomial, we must have $\alpha_1 = \cdots = \alpha_m = 0$. Thus, the list $f_1, \ldots, f_m$ is linearly independent. Therefore, by Exercise 2A17, the real vector space of all continuous real-valued functions on the interval $[0, 1]$ is infinite-dimensional. □

Solution 2A.20

Appendix. *An alternate proof is given at the end of this solution.*

Note that $\mathbb{F}[X]_{\leq m}$ is spanned by the list $1, z, \ldots, z^m$ of $m + 1$ vectors, so there cannot be a list of $m + 2$ linearly independent vectors in $\mathbb{F}[X]_{\leq m}$. Suppose $p_0, \ldots, p_m$ is linearly independent in $\mathbb{F}[X]_{\leq m}$. Moreover, notice that every linear combination of $p_0, \ldots, p_m$ is a polynomial such that

$$\sum_{i=0}^{m} \alpha_i p_i(2) = 0.$$

Since $q = 1 \in \mathbb{F}[X]_{\leq m}$ is a polynomial where $q(2) \neq 0$, $q \notin \operatorname{span}(p_0, \ldots, p_n)$. By Exercise 2A13, the list $p_0, \ldots, p_m, q$ is of length $m + 2$ that is linearly independent, which is a contradiction. Hence, $p_0, \ldots, p_m$ cannot be linearly independent in $\mathbb{F}[X]_{\leq m}$.

Alternate proof. Since $p_k(2) = 0$ for each $k \in \{0, \ldots, m\}$, we have

$$p_k(x) = (x - 2)q_k(x)$$

for some polynomial $q_k(x) \in \mathbb{F}[X]_{\leq m-1}$. Then we may write

$$p_0(x), p_1(x), \ldots, p_m(x) = (x - 2)q_0(x), (x - 2)q_1(x), \ldots, (x - 2)q_m(x).$$

Since $\mathbb{F}[X]_{\leq m-1}$ can be spanned by the m polynomials $1, x, \ldots, x^{m-1}$, the list of $m + 1$ polynomials $q_0, q_1, \ldots, q_m$ must be linearly dependent. Thus, there exist scalars $\alpha_0, \ldots, \alpha_m$, not all zero, such that

$$\alpha_0 q_0(x) + \cdots + \alpha_m q_m(x) = 0.$$

Then we may write

$$\alpha_0 p_0(x) + \cdots + \alpha_m p_m(x) = (x - 2)(\alpha_0 q_0(x) + \cdots + \alpha_m q_m(x)) = 0.$$

Since not all the scalars $\alpha_0, \ldots, \alpha_m$ are zero, this shows that $p_0, p_1, \ldots, p_m$ is linearly dependent in $\mathbb{F}[X]_{\leq m}$. □

Solution 2B.1

Manual Remark. *The proof of this exercise fails if* $\mathbb{F}$ *contains only two elements. In this case, an additional vector space with exactly one basis is the vector space of dimension one over* $\mathbb{F}$*, since the only nonzero scalar is* 1*, and any nonzero vector* v *in the vector space is a basis.*

Let V be a vector space with exactly one basis, call it $v_1, \ldots, v_n$. Consider that for $v \in V$, the list containing only v is linearly independent, so we may extend the list v to a basis for V. This means that $v = v_k$ for some $k \in \{1, \ldots, n\}$, so every vector in V is one of $v_1, \ldots, v_n$. For nonzero $\lambda \in \mathbb{F}$ and $k \in \{1, \ldots, n\}$, consider $\lambda v_k \in V$. If $\lambda v_k = v_j$ for some $j \neq k$, then $v_1, \ldots, v_n$ is not linearly independent, a contradiction. Therefore, $\lambda v_k = v_k$, and thus $(\lambda - 1)v_k = 0$. Because λ was arbitrary, it must be that $v_k = 0$. Because k was arbitrary, we have $v_k = 0$ for all $k \in \{1, \ldots, n\}$, so the basis for V is empty. span( ) $= \{0\}$, so the trivial vector space $V = \{0\}$ is the only vector space with exactly one basis. □

Solution 2B.2

(a) Suppose $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$\alpha_1(1, 0, \ldots, 0) + \alpha_2(0, 1, 0, \ldots, 0) + \cdots + \alpha_n(0, \ldots, 0, 1) = 0.$$

Then each $\alpha_i = 0$, so the list is linearly independent. Moreover, any vector $(x_1, \ldots, x_n) \in \mathbb{F}^n$ can be written as

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

so the list spans $\mathbb{F}^n$. Thus, the list is a basis of $\mathbb{F}^n$.

(b) If $(3, 5) = a(1, 2)$ for some $a \in \mathbb{F}$, then we must have $3 = a$ and $5 = 2a$, which is impossible. Thus, the list is linearly independent. Moreover, for any element $(x, y) \in \mathbb{F}^2$, we can solve the system of equations

$$\begin{aligned} x &= 3\alpha + \beta, \\ y &= 5\alpha + 2\beta \end{aligned}$$

for $\alpha, \beta \in \mathbb{F}$. Solving this system of equations, we find that

$$\alpha = 2x - y, \quad \beta = -5x + 3y.$$

Since we can find $\alpha, \beta \in \mathbb{F}$ for any $(x, y) \in \mathbb{F}^2$, the list spans $\mathbb{F}^2$. Thus, the list is a basis of $\mathbb{F}^2$.

(c) Note that $\mathbb{F}^3$ is spanned by the linearly independent list $(1, 0, 0), (0, 1, 0), (0, 0, 1)$. Since the length of a spanning list must be at least the length of a linearly independent list, the list $(1, 2, -4), (7, -5, 6)$ cannot span $\mathbb{F}^3$.

(d) From part (b), we saw that $(1, 2), (3, 5)$ is a basis of $\mathbb{F}^2$. Since $(4, 13) = -5(3, 5) + 19(1, 2)$, the list $(1, 2), (3, 5), (4, 13)$ is linearly dependent. Moreover, since $(1, 2), (3, 5)$ is a basis of $\mathbb{F}^2$, the list $(1, 2), (3, 5), (4, 13)$ spans $\mathbb{F}^2$. Thus, the list is not a basis of $\mathbb{F}^2$.

(e) Let U be the subspace in question. If $\alpha, \beta \in \mathbb{F}$ such that

$$\alpha(1,1,0) + \beta(0,0,1) = (0,0,0),$$

then we must have $\alpha = \beta = 0$, so the list is linearly independent. Moreover, for any $(x,x,y) \in U$, we can write

$$(x,x,y) = x(1,1,0) + y(0,0,1),$$

so the list spans U. Thus, the list is a basis of U.

(f) Let U be the subspace in question. If $\alpha, \beta \in \mathbb{F}$ such that

$$\alpha(1,-1,0) + \beta(1,0,-1) = (0,0,0),$$

then we must have $\alpha = \beta = 0$, so the list is linearly independent. Moreover, for any $(x,y,z) \in U$, note that $x + y + z = 0$, implying that $x = -y - z$. Then we can write

$$(x,y,z) = (-y-z, y, z) = -y(1,-1,0) - z(1,0,-1),$$

so the list spans U. Thus, the list is a basis of U.

(g) If $\alpha_0, \ldots, \alpha_m \in \mathbb{F}$ such that

$$\alpha_0 + \alpha_1 z + \cdots + \alpha_m z^m = 0,$$

then we must have $\alpha_0 = \cdots = \alpha_m = 0$, so the list is linearly independent. Moreover, any polynomial $p(z) \in \mathbb{F}[X]_{\leq m}$ can be written as

$$p(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_m z^m,$$

so the list spans $\mathbb{F}[X]_{\leq m}$. Thus, the list is a basis of $\mathbb{F}[X]_{\leq m}$.

□

Solution 2B.3

(a) Note that we can rewrite the subspace U as

$$U = \{(3x_2, x_2, 7x_4, x_4, x_5) \in \mathbb{R}^5 : x_2, x_4, x_5 \in \mathbb{R}\}.$$

Then we can write any vector in U as a linear combination of the following three vectors:

$$u_1 = (3,1,0,0,0), \quad u_2 = (0,0,7,1,0), \quad u_3 = (0,0,0,0,1).$$

Moreover, these three vectors are linearly independent, since the only way to write the zero vector as a linear combination of u_1, u_2, u_3 is to take all coefficients to be zero. Thus, $\{u_1, u_2, u_3\}$ is a basis of U.

(b) We can extend the basis $\{u_1, u_2, u_3\}$ of U to a basis of $\mathbb{R}^5$ by adding two more vectors that are linearly independent of u_1, u_2, u_3. For example, we can add the vectors

$$u_4 = (1,0,0,0,0), \quad u_5 = (0,0,1,0,0),$$

which are linearly independent of u_1, u_2, u_3: since u_1 is the only vector in the list with a nonzero first component, u_4 cannot be a linear combination of u_1, u_2, u_3. Similarly, since u_2 is the only vector in the list with a nonzero third component, u_5 cannot be a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{R}^5$.

(c) Let W be the subspace of $\mathbb{R}^5$ spanned by the vectors u_4 and u_5. Then we have

$$\mathbb{R}^5 = U \oplus W,$$

since $U \cap W = \{0\}$ and every vector in $\mathbb{R}^5$ can be written as a sum of a vector in U and a vector in W.

□

Solution 2B.4

(a) Note that we can rewrite the subspace U as

$$U = \{(z_1, 6z_1, -2z_4 - 3z_5, z_4, z_5) \in \mathbb{C}^5 : z_1, z_4, z_5 \in \mathbb{C}\}.$$

Then we can write any vector in U as a linear combination of the following three vectors:

$$u_1 = (1, 6, 0, 0, 0), \quad u_2 = (0, 0, -2, 1, 0), \quad u_3 = (0, 0, -3, 0, 1).$$

Moreover, these three vectors are linearly independent, since the only way to write the zero vector as a linear combination of u_1, u_2, u_3 is to take all coefficients to be zero. Thus, $\{u_1, u_2, u_3\}$ is a basis of U.

(b) We can extend the basis $\{u_1, u_2, u_3\}$ of U to a basis of $\mathbb{C}^5$ by adding two more vectors that are linearly independent of u_1, u_2, u_3. For example, we can add the vectors

$$u_4 = (1, 0, 0, 0, 0), \quad u_5 = (0, 0, 1, 0, 0),$$

which are linearly independent of u_1, u_2, u_3: since u_1 is the only vector in the list with a nonzero first component, u_4 cannot be a linear combination of u_1, u_2, u_3. Similarly, u_2 and u_3 have nonzero third components, but any linear combination of u_2 and u_3 will have nonzero fourth and fifth components, so u_5 cannot be a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{C}^5$.

(c) Let W be the subspace of $\mathbb{C}^5$ spanned by the vectors u_4 and u_5. Then we have

$$\mathbb{C}^5 = U \oplus W,$$

since $U \cap W = \{0\}$ and every vector in $\mathbb{C}^5$ can be written as a sum of a vector in U and a vector in W.

□

Solution 2B.5 Let $u_1, \ldots, u_m$ be a basis of U and $w_1, \ldots, w_n$ a basis of W. Then the list $u_1, \ldots, u_m, w_1, \ldots, w_n$ spans V, since any vector in V can be written as a sum of a vector in U and a vector in W. Because every spanning list of V can be reduced to a basis of V, $u_1, \ldots, u_m, w_1, \ldots, w_n$ can be reduced to a basis of V consisting of vectors in $U \cup W$. □

Solution 2B.6 This is not true. Let $p_0 = 1, p_1 = x, p_2 = x^3, p_3 = x^2 + x^3$. Then none of the polynomials p_0, p_1, p_2, p_3 has degree 2. Moreover, the list p_0, p_1, p_2, p_3 is linearly independent, since the only way to write the zero polynomial as a linear combination of p_0, p_1, p_2, p_3 is to take all coefficients to be zero. Finally, noting that $x^2 = p_3 - p_2$, then for any polynomial $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 \in \mathbb{F}[X]_{\leq 3}$, we can write

$$p(x) = a_0p_0 + a_1p_1 + a_2(p_3 - p_2) + a_3p_2 = a_0p_0 + a_1p_1 + (a_3 - a_2)p_2 + a_2p_3,$$

which shows that the list spans $\mathbb{F}[X]_{\leq 3}$. Thus, the list p_0, p_1, p_2, p_3 is a basis of $\mathbb{F}[X]_{\leq 3}$. □

Solution 2B.7

Manual Remark. *The technique of transforming the scalars $\beta_1, \beta_2, \beta_3, \beta_4$ into the scalars $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ is good to remember. You will later learn that this is called a* change of basis.

Let $\alpha_1, \ldots, \alpha_4 \in \mathbb{F}$ such that

$$\alpha_1(v_1 + v_2) + \alpha_2(v_2 + v_3) + \alpha_3(v_3 + v_4) + \alpha_4v_4 = 0.$$

Then we may write

$$\alpha_1v_1 + (\alpha_1 + \alpha_2)v_2 + (\alpha_2 + \alpha_3)v_3 + (\alpha_3 + \alpha_4)v_4 = 0.$$

Since v_1, v_2, v_3, v_4 is a basis of V, we must have

$$\alpha_1 = 0, \quad \alpha_1 + \alpha_2 = 0, \quad \alpha_2 + \alpha_3 = 0, \quad \alpha_3 + \alpha_4 = 0.$$

From the first equation, we have $\alpha_1 = 0$. Substituting this into the second equation gives us $\alpha_2 = 0$. Substituting this into the third equation gives us $\alpha_3 = 0$. Finally, substituting this into the fourth equation gives us $\alpha_4 = 0$. Thus, the list is linearly independent. Moreover, since v_1, v_2, v_3, v_4 is a basis of V, then any vector $v \in V$ can be written as

$$v = \beta_1v_1 + \beta_2v_2 + \beta_3v_3 + \beta_4v_4$$

for some $\beta_1, \ldots, \beta_4 \in \mathbb{F}$. Using the expressions from proving the linear independence of the list, we may show that there exist $\alpha_1, \ldots, \alpha_4 \in \mathbb{F}$ such that

$$\begin{cases} \beta_1 = \alpha_1, \\ \beta_2 = \alpha_1 + \alpha_2, \\ \beta_3 = \alpha_2 + \alpha_3, \\ \beta_4 = \alpha_3 + \alpha_4. \end{cases}$$

Then $v \in V$ may be written as

$$v = \alpha_1(v_1 + v_2) + \alpha_2(v_2 + v_3) + \alpha_3(v_3 + v_4) + \alpha_4v_4,$$

where

$$\alpha_1 = \beta_1, \quad \alpha_2 = \beta_2 - \beta_1, \quad \alpha_3 = \beta_3 - \beta_2 + \beta_1, \quad \alpha_4 = \beta_4 - \beta_3 + \beta_2 - \beta_1.$$

Thus, the list spans V. Therefore, the list is a basis of V. □

Solution 2B.8 This is not true. Let $V = \mathbb{R}^4$ and v_1, v_2, v_3, v_4 be the standard basis of $\mathbb{R}^4$. Let U be the subspace of $\mathbb{R}^4$ defined by

$$U = \operatorname{span}(v_1, v_2, v_3 + v_4) = \{(x, y, z, z) \in \mathbb{R}^4 : x, y, z \in \mathbb{R}\}.$$

Then $v_1, v_2 \in U$ and $v_3, v_4 \notin U$. However, v_1, v_2 does not span U, since $(0, 0, 1, 1) \in U$ but cannot be written as a linear combination of v_1 and v_2. Thus, v_1, v_2 is not a basis of U. □

Solution 2B.9 By Exercise 2A14, we know that $v_1, \ldots, v_m$ is linearly independent if and only if $w_1, \ldots, w_m$ is linearly independent. By Exercise 2A3, we know that $v_1, \ldots, v_m$ spans V if and only if $w_1, \ldots, w_m$ spans V. Thus, $v_1, \ldots, v_m$ is a basis of V if and only if $w_1, \ldots, w_m$ is a basis of V. □

Solution 2B.10 Let $v \in V$. Since $V = U \oplus W$, there are unique $u \in U$ and $w \in W$ such that $v = u + w$. Since $u_1, \ldots, u_m$ spans U and $w_1, \ldots, w_n$ spans W, there exists scalars $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_n$ such that

$$u = \sum_{i=1}^{m} \alpha_i u_i \quad \text{and} \quad w = \sum_{j=1}^{n} \beta_j w_j.$$

Therefore,

$$v = u + w = \sum_{i=1}^{m} \alpha_i u_i + \sum_{j=1}^{n} \beta_j w_j,$$

which shows that the list spans V. To show linear independence, suppose $v = 0$ so that $v = u + w$ imply $u = w = 0$. Since $u_1, \ldots, u_m$ and $w_1, \ldots, w_n$ are linearly independent, it follows that $\alpha_1 = \cdots = \alpha_m = 0$ and $\beta_1 = \cdots = \beta_n = 0$. Therefore, the list $u_1, \ldots, u_m, w_1, \ldots, w_n$ is linearly independent. Combining this with the fact that it spans V, we conclude that it is a basis of V. □

Solution 2B.11

Axler Note. *Please see Exercise 1B8 for the definition of the complexification* $V_{\mathbb{C}}$

Let $v_1, \ldots, v_n$ be a basis of V. We claim that $v_1, \ldots, v_n$ is also a basis of $V_{\mathbb{C}}$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$ such that

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0.$$

Then we may write each $\alpha_k = a_k + b_k i$ for some $a_k, b_k \in \mathbb{R}$. Then we have

$$(a_1 + b_1 i)v_1 + \cdots + (a_n + b_n i)v_n = 0.$$

This implies that

$$(a_1 v_1 + \cdots + a_n v_n) + i(b_1 v_1 + \cdots + b_n v_n) = 0.$$

Since the left-hand side is a sum of two vectors in V, and V is a real vector space, we must have

$$a_1 v_1 + \cdots + a_n v_n = 0 \quad \text{and} \quad b_1 v_1 + \cdots + b_n v_n = 0.$$

Since $v_1, \ldots, v_n$ is a basis of V, we have $a_1 = \cdots = a_n = 0$ and $b_1 = \cdots = b_n = 0$. Thus, $\alpha_1 = \cdots = \alpha_n = 0$, which shows that the list is linearly independent. Moreover, let $v \in V_{\mathbb{C}}$. Then we may write $v = u + wi$ for some $u, w \in V$. Since $v_1, \ldots, v_n$ is a basis of V, we may write

$$u = \gamma_1 v_1 + \cdots + \gamma_n v_n \quad \text{and} \quad w = \delta_1 v_1 + \cdots + \delta_n v_n$$

for some $\gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_n \in \mathbb{R}$. Hence,

$$v = u + wi = (\gamma_1 + \delta_1 i)v_1 + \cdots + (\gamma_n + \delta_n i)v_n,$$

which shows that the list spans $V_{\mathbb{C}}$. Therefore, the list is a basis of $V_{\mathbb{C}}$. □

§2C

Solution 2C.1 Since $\dim(\mathbb{R}^2) = 2$, any subspace of $\mathbb{R}^2$ must have dimension 0, 1, or 2. The only subspace of dimension 0 is $\{0\}$, and the only subspace of dimension 2 is $\mathbb{R}^2$. Any subspace with dimension 1 consists of all scalar multiples of some nonzero $(x, y) \in \mathbb{R}^2$, which is precisely the line with direction (x, y) containing the origin. Thus, the subspaces of $\mathbb{R}^2$ are precisely $\{0\}$, all lines in $\mathbb{R}^2$ containing the origin, and $\mathbb{R}^2$. □

Solution 2C.2 Since $\dim(\mathbb{R}^3) = 3$, any subspace of $\mathbb{R}^3$ must have dimension 0, 1, 2, or 3. The only subspace of dimension 0 is $\{0\}$, and the only subspace of dimension 3 is $\mathbb{R}^3$. Any subspace with dimension 1 consists of all scalar multiples of some nonzero $(x, y, z) \in \mathbb{R}^3$, which is precisely the line containing the origin in the direction of (x, y, z). Any subspace with dimension 2 consists of all linear combinations of two linearly independent vectors in $\mathbb{R}^3$, which is precisely a plane containing the origin. Thus, the subspaces of $\mathbb{R}^3$ are precisely $\{0\}$, all lines in $\mathbb{R}^3$ containing the origin, all planes in $\mathbb{R}^3$ containing the origin, and $\mathbb{R}^3$. □

Solution 2C.3

Manual Remark. *Please see remark at the beginning of this section regarding the basis of U.*

(a) Let $p = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 \in U$. Then we have $p(6) = 0$, which gives us the equation

$$a_0 + 6a_1 + 36a_2 + 216a_3 + 1296a_4 = 0.$$

We can solve this equation for a_0 in terms of a_1, a_2, a_3, a_4:

$$a_0 = -6a_1 - 36a_2 - 216a_3 - 1296a_4.$$

Then we can write any polynomial in U as a linear combination of the following four polynomials:

$$u_1 = -6 + x, \quad u_2 = -36 + x^2, \quad u_3 = -216 + x^3, \quad u_4 = -1296 + x^4.$$

(b) We can extend the basis $\{u_1, u_2, u_3, u_4\}$ of U to a basis of $\mathbb{F}[X]_{\leq 4}$ by adding one more polynomial that is linearly independent of u_1, u_2, u_3, u_4. For example, we can add the polynomial

$$u_5 = 1,$$

which is linearly independent of u_1, u_2, u_3, u_4: since u_5 has no higher-degree terms, it cannot be expressed as a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{F}[X]_{\leq 4}$.

(c) Let $W = \operatorname{span} u_5 = \operatorname{span} 1$. Then we have

$$\mathbb{F}[X]_{\leq 4} = U + W,$$

since every polynomial in $\mathbb{F}[X]_{\leq 4}$ can be written as a sum of a polynomial in U and a polynomial in W. Moreover, if $p \in U \cap W$, then $p = \alpha$ for some $\alpha \in \mathbb{F}$. Since $p(6) = 0$, we have $\alpha = 0$. Thus, $U \cap W = \{0\}$, and so we have

$$\mathbb{F}[X]_{\leq 4} = U \oplus W. \qquad \square$$

Solution 2C.4

Manual Remark. *Please see remark at the beginning of this section regarding the basis of U.*

(a) Let $p = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 \in U$. Then we have

$$p''(x) = 2a_2 + 6a_3 x + 12a_4 x^2,$$

and so $p''(6) = 0$ gives us the equation

$$2a_2 + 36a_3 + 432a_4 = 0.$$

We can solve this equation for a_2 in terms of a_3, a_4:

$$a_2 = -18a_3 - 216a_4.$$

Then we can write any polynomial in U as a linear combination of the following four polynomials:

$$u_1 = 1, \quad u_2 = x, \quad u_3 = -18x^2 + x^3, \quad u_4 = -216x^2 + x^4.$$

(b) We can extend the basis $\{u_1, u_2, u_3, u_4\}$ of U to a basis of $\mathbb{R}[X]_{\leq 4}$ by adding one more polynomial that is linearly independent of u_1, u_2, u_3, u_4. For example, we can add the polynomial

$$u_5 = x^2,$$

which is linearly independent of u_1, u_2, u_3, u_4: since only u_3 and u_4 have x^2 terms but contain other higher-degree terms, u_5 cannot be expressed as a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{R}[X]_{\leq 4}$.

(c) Let $W = \operatorname{span} u_5 = \operatorname{span} x^2$. Then we have

$$\mathbb{R}[X]_{\leq 4} = U + W,$$

since every polynomial in $\mathbb{R}[X]_{\leq 4}$ can be written as a sum of a polynomial in U and a polynomial in W. Moreover, if $p \in U \cap W$, then $p = \alpha x^2$ for some $\alpha \in \mathbb{R}$. Since $p''(6) = 0$, we have $2\alpha = 0$, which implies that $\alpha = 0$. Thus, $U \cap W = \{0\}$, and so we have

$$\mathbb{R}[X]_{\leq 4} = U \oplus W. \qquad \square$$

Solution 2C.5

Manual Remark. *Please see remark at the beginning of this section regarding the basis of U.*

(a) Let $p = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 \in U$. Then we have $p(2) = p(5)$, which gives us the equation

$$a_0 + 2a_1 + 4a_2 + 8a_3 + 16a_4 = a_0 + 5a_1 + 25a_2 + 125a_3 + 625a_4.$$

We can simplify this equation to get

$$-3a_1 - 21a_2 - 117a_3 - 609a_4 = 0.$$

We can solve this equation for a_1 in terms of a_2, a_3, a_4:

$$a_1 = -7a_2 - 39a_3 - 203a_4.$$

Then we can write any polynomial in U as a linear combination of the following four polynomials:

$$u_1 = 1, \quad u_2 = -7x + x^2, \quad u_3 = -39x + x^3, \quad u_4 = -203x + x^4.$$

(b) We can extend the basis $\{u_1, u_2, u_3, u_4\}$ of U to a basis of $\mathbb{F}[X]_{\leq 4}$ by adding one more polynomial that is linearly independent of u_1, u_2, u_3, u_4. For example, we can add the polynomial

$$u_5 = x,$$

which is linearly independent of u_1, u_2, u_3, u_4: since only u_2, u_3, u_4 have x terms but contain other higher-degree terms, u_5 cannot be expressed as a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{F}[X]_{\leq 4}$.

(c) Let $W = \operatorname{span} u_5 = \operatorname{span} x$. Then we have

$$\mathbb{F}[X]_{\leq 4} = U + W,$$

since every polynomial in $\mathbb{F}[X]_{\leq 4}$ can be written as a sum of a polynomial in U and a polynomial in W. Moreover, if $p \in U \cap W$, then $p = \alpha x$ for some $\alpha \in \mathbb{F}$. Since $p(2) = p(5)$, we have $2\alpha = 5\alpha$, which implies that $\alpha = 0$. Thus, $U \cap W = \{0\}$, and so we have

$$\mathbb{F}[X]_{\leq 4} = U \oplus W. \qquad \square$$

Solution 2C.6

Manual Remark. *Please see remark at the beginning of this section regarding the basis of U.*

(a) Let $p = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 \in U$. Then we have $p(2) = p(5) = p(6)$. We know from Exercise 2C5 that the condition $p(2) = p(5)$ gives us the equation

$$a_1 = -7a_2 - 39a_3 - 203a_4.$$

The condition $p(5) = p(6)$ gives us the equation

$$5a_1 + 25a_2 + 125a_3 + 625a_4 = 6a_1 + 36a_2 + 216a_3 + 1296a_4,$$

which simplifies to

$$a_1 + 11a_2 + 91a_3 + 671a_4 = 0.$$

Since we know that $a_1 = -7a_2 - 39a_3 - 203a_4$, we can substitute this into the second equation to get

$$4a_2 + 52a_3 + 468a_4 = 0.$$

We can solve this equation for a_2 in terms of a_3, a_4:

$$a_2 = -13a_3 - 117a_4.$$

Then $a_1 = 52a_3 + 616a_4$. Thus, we can write any polynomial in U as a linear combination of the following three polynomials:

$$u_1 = 1, \quad u_2 = 52x - 13x^2 + x^3, \quad u_3 = 616x - 117x^2 + x^4.$$

(b) We can extend the basis $\{u_1, u_2, u_3\}$ of U to a basis of $\mathbb{F}[X]_{\leq 4}$ by adding two more polynomials that are linearly independent of u_1, u_2, u_3. For example, we can add the polynomials

$$u_4 = x^2, \quad u_5 = x^3,$$

which are linearly independent of u_1, u_2, u_3: since only u_2 and u_3 have x^2 terms but contain other higher-degree terms, u_4 cannot be expressed as a linear combination of u_1, u_2, u_3. Similarly, u_5 has an x^3 term but no higher-degree terms, so it cannot be expressed as a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{F}[X]_{\leq 4}$.

(c) Let $W = \operatorname{span}(u_4, u_5) = \operatorname{span}(x^2, x^3)$. Then we have

$$\mathbb{F}[X]_{\leq 4} = U + W,$$

since every polynomial in $\mathbb{F}[X]_{\leq 4}$ can be written as a sum of a polynomial in U and a polynomial in W. Moreover, if $p \in U \cap W$, then $p = \alpha x^2 + \beta x^3$ for some $\alpha, \beta \in \mathbb{F}$. Since $p(2) = p(5) = p(6)$, we have

$$4\alpha + 8\beta = 25\alpha + 125\beta = 36\alpha + 216\beta.$$

Solving this system of equations, we find that $\alpha = \beta = 0$. Thus, $U \cap W = \{0\}$, and so we have

$$\mathbb{F}[X]_{\leq 4} = U \oplus W. \qquad \square$$

Solution 2C.7

Manual Remark. *Please see remark at the beginning of this section regarding the basis of U.*

(a) Let $p = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 \in U$. Then we have

$$\begin{aligned}\int_{-1}^{1} p &= \int_{-1}^{1} (a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4)\,dx \\ &= \int_{-1}^{1} (a_0 + a_2x^2 + a_4x^4)\,dx \\ &= 2\int_{0}^{1} (a_0 + a_2x^2 + a_4x^4)\,dx \\ &= 2\left(a_0 + \frac{a_2}{3} + \frac{a_4}{5}\right) = 0.\end{aligned}$$

We can solve this equation for a_0 in terms of a_2, a_4:

$$a_0 = -\frac{a_2}{3} - \frac{a_4}{5}.$$

Then we can write any polynomial in U as a linear combination of the following four polynomials:

$$u_1 = x, \quad u_2 = x^2 - \frac{1}{3}, \quad u_3 = x^3, \quad u_4 = x^4 - \frac{1}{5}.$$

(b) We can extend the basis $\{u_1, u_2, u_3, u_4\}$ of U to a basis of $\mathbb{R}[X]_{\leq 4}$ by adding one more polynomial that is linearly independent of u_1, u_2, u_3, u_4. For example, we can add the polynomial

$$u_5 = 1,$$

which is linearly independent of u_1, u_2, u_3, u_4: since u_5 has no higher-degree terms, it cannot be expressed as a linear combination of u_1, u_2, u_3, u_4. Thus, $\{u_1, u_2, u_3, u_4, u_5\}$ is a basis of $\mathbb{R}[X]_{\leq 4}$.

(c) Let $W = \operatorname{span} u_5 = \operatorname{span} 1$. Then we have

$$\mathbb{R}[X]_{\leq 4} = U + W,$$

since every polynomial in $\mathbb{R}[X]_{\leq 4}$ can be written as a sum of a polynomial in U and a polynomial in W. Moreover, if $p \in U \cap W$, then $p = \alpha$ for some $\alpha \in \mathbb{R}$. Since the integral of a constant over a symmetric interval is nonzero unless the function is identically zero, we have $\alpha = 0$. Thus, $U \cap W = \{0\}$, and so we have

$$\mathbb{R}[X]_{\leq 4} = U \oplus W. \qquad \square$$

Solution 2C.8 Consider the subspaces of V:

$$U = \operatorname{span}(v_1, \ldots, v_m), \quad X = \operatorname{span}(v_1 + w, \ldots, v_m + w), \quad W = \operatorname{span}(w).$$

Since $v_1, \ldots, v_m$ is linearly independent, we have $\dim U = m$. Moreover, $\dim W = 1$. Lastly, note that $v_k = (v_k + w) - w \in X + W$ for each $k = 1, \ldots, m$. Therefore, $U \subseteq X + W$, and since U is a subspace of dimension m, we have

$$\dim U \leq \dim(X + W) = \dim X + \dim W - \dim(X \cap W) \leq \dim X + \dim W,$$

which gives

$$m = \dim U \leq \dim X + \dim W = \dim X + 1 \implies \dim X \geq m - 1. \qquad \square$$

Solution 2C.9 Suppose that

$$\alpha_0 p_0 + \alpha_1 p_1 + \cdots + \alpha_m p_m = 0$$

for some $\alpha_0, \ldots, \alpha_m \in \mathbb{F}$. Since p_m is the only polynomial with nonzero coefficient of x^m, we must have $\alpha_m = 0$. Next, since p_{m-1} is the only polynomial with nonzero coefficient of x^{m-1}, we must have $\alpha_{m-1} = 0$. Continuing in this way, we find that $\alpha_0 = \cdots = \alpha_m = 0$. Thus, the list is linearly independent. Moreover, since $\dim \mathbb{F}[X]_{\leq m} = m + 1$, any linearly independent list of $m + 1$ vectors in $\mathbb{F}[X]_{\leq m}$ is a basis. Therefore, $p_0, p_1, \ldots, p_m$ is a basis of $\mathbb{F}[X]_{\leq m}$. $\square$

Solution 2C.10 Suppose that

$$\alpha_0 p_0 + \alpha_1 p_1 + \cdots + \alpha_m p_m = 0$$

for some $\alpha_0, \ldots, \alpha_m \in \mathbb{F}$. Since p_0 is the only polynomial with nonzero constant term, we must have $\alpha_0 = 0$, hence the sum above becomes

$$\alpha_1 p_1 + \cdots + \alpha_m p_m = 0.$$

Next, since p_1 is the only polynomial with nonzero coefficient of x in the sum above, we must have $\alpha_1 = 0$. Continuing in this way, we find that $\alpha_0 = \cdots = \alpha_m = 0$. Thus, the list is linearly independent. Moreover, since $\dim \mathbb{F}[X]_{\leq m} = m + 1$, any linearly independent list of $m + 1$ vectors in $\mathbb{F}[X]_{\leq m}$ is a basis. Therefore, $p_0, \ldots, p_m$ is a basis of $\mathbb{F}[X]_{\leq m}$. $\square$

Solution 2C.11 Let U and W be four-dimensional subspaces of $\mathbb{C}^6$. By the dimension formula for the sum of two subspaces, we have

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W) = 4 + 4 - \dim(U \cap W) = 8 - \dim(U \cap W).$$

Since $U + W$ is a subspace of $\mathbb{C}^6$, we have $\dim(U + W) \leq 6$. Thus, we have

$$8 - \dim(U \cap W) \leq 6,$$

which implies that $\dim(U \cap W) \geq 2$. Therefore, there exist at least two linearly independent vectors in $U \cap W$, and neither of these vectors is a scalar multiple of the other. $\square$

Solution 2C.12 By the dimension formula for the sum of two subspaces, we have

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W) = 3+5-\dim(U \cap W) = 8-\dim(U \cap W).$$

Since $U + W = \mathbb{R}^8$, we have $\dim(U + W) = 8$. Thus, we have

$$8 - \dim(U \cap W) = 8,$$

which implies that $\dim(U \cap W) = 0$. Therefore, $U \cap W = \{0\}$, and so we have

$$\mathbb{R}^8 = U \oplus W. \qquad \square$$

Solution 2C.13 By the dimension formula for the sum of two subspaces, we have

$$\dim(U+W) = \dim U + \dim W - \dim(U \cap W) = 5+5-\dim(U \cap W) = 10-\dim(U \cap W).$$

Since $U + W$ is a subspace of $\mathbb{R}^8$, we have $\dim(U + W) \leq 8$. Thus, we have

$$10 - \dim(U \cap W) \leq 8,$$

which implies that $\dim(U \cap W) \geq 2$. Therefore, $U \cap W$ is nontrivial, and so we have $U \cap W \neq \{0\}$. $\square$

Solution 2C.14 Let $W = V_1 \cap V_2$. By the dimension formula for the sum of two subspaces, we have

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim W = 7 + 7 - \dim W = 14 - \dim W.$$

Since $V_1 + V_2$ is a subspace of V, we have $\dim(V_1 + V_2) \leq 10$. Thus, we have $14 - \dim W \leq 10$, which implies that $\dim W \geq 4$. Therefore, $V_1 \cap V_2$ is at least four-dimensional. Now, consider the subspace $W \cap V_3 = V_1 \cap V_2 \cap V_3$. By the dimension formula for the sum of two subspaces, we have

$$\dim(W + V_3) = \dim W + \dim V_3 - \dim(W \cap V_3) = \dim W + 7 - \dim(W \cap V_3).$$

Since $W + V_3$ is a subspace of V, we have $\dim(W + V_3) \leq 10$. Thus, we have

$$\dim W + 7 - \dim(W \cap V_3) \leq 10,$$

which implies that $\dim(W \cap V_3) \geq \dim W + 7 - 10 = \dim W - 3$. Since $\dim W \geq 4$, we have $\dim(W \cap V_3) \geq 1$. Therefore, $V_1 \cap V_2 \cap V_3$ is nontrivial, and so we have $V_1 \cap V_2 \cap V_3 \neq \{0\}$. $\square$

Solution 2C.15 Let $W = V_1 \cap V_2$. By the dimension formula for the sum of two subspaces, we have

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim W.$$

Since $V_1 + V_2$ is a subspace of V, we have $\dim(V_1 + V_2) \leq \dim V$. Thus, we have

$$\dim V_1 + \dim V_2 - \dim W \leq \dim V,$$

which implies that $\dim W \geq \dim V_1 + \dim V_2 - \dim V$. Now, consider the subspace $W \cap V_3 = V_1 \cap V_2 \cap V_3$. By the dimension formula for the sum of two subspaces, we have

$$\dim(W + V_3) = \dim W + \dim V_3 - \dim(W \cap V_3).$$

Since $W + V_3$ is a subspace of V, we have $\dim(W + V_3) \leq \dim V$. Thus, we have

$$\dim W + \dim V_3 - \dim(W \cap V_3) \leq \dim V,$$

which implies that $\dim(W \cap V_3) \geq \dim W + \dim V_3 - \dim V$. Since $\dim W \geq \dim V_1 + \dim V_2 - \dim V$, we have

$$\dim(W \cap V_3) \geq (\dim V_1 + \dim V_2 - \dim V) + \dim V_3 - \dim V = \dim V_1 + \dim V_2 + \dim V_3 - 2\dim V.$$

Since $\dim V_1 + \dim V_2 + \dim V_3 > 2\dim V$, we have $\dim(W \cap V_3) > 0$. Therefore, $V_1 \cap V_2 \cap V_3$ is nontrivial, and so we have $V_1 \cap V_2 \cap V_3 \neq \{0\}$. □

Solution 2C.16 Let U be a subspace of V with $\dim U = m < n = \dim V$. We can extend a basis of U to a basis of V. Let $\{u_1, \ldots, u_m\}$ be a basis of U, and extend it to a basis of V by adding vectors $v_1, \ldots, v_{n-m}$. Then $\{u_1, \ldots, u_m, v_1, \ldots, v_{n-m}\}$ is a basis of V. For each $i = 1, \ldots, n - m$, let

$$W_i = \operatorname{span}(u_1, \ldots, u_m, v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{n-m}).$$

By construction, W_i does not contain the extended basis vector v_i. Moreover, each W_i is a subspace of V with dimension $n - 1$, since it is spanned by $n - 1$ linearly independent vectors. Additionally, we have

$$U = W_1 \cap W_2 \cap \cdots \cap W_{n-m},$$

since any vector in the intersection containing nonzero v_i would not be in W_i, leaving behind only the basis vectors in U. Therefore, there exist $n - m$ subspaces of V, each of dimension $n - 1$, whose intersection equals U. □

Solution 2C.17 We will prove the statement by induction on m. For the base case, let $m = 2$. Then the sum $V_1 + V_2$ is finite-dimensional. By the dimension formula for the sum of two subspaces, we have

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2) \leq \dim V_1 + \dim V_2.$$

Hence, the statement holds for $m = 2$. Now, suppose that the statement holds for some positive integer m. That is, suppose that if $V_1, \ldots, V_m$ are finite-dimensional subspaces of V, then $V_1 + \cdots + V_m$ is finite-dimensional and

$$\dim(V_1 + \cdots + V_m) \leq \dim V_1 + \cdots + \dim V_m.$$

We want to show that the statement holds for $m + 1$. Let $V_1, \ldots, V_{m+1}$ be finite-dimensional subspaces of V. By the induction hypothesis, we know that $W = V_1 + \cdots + V_m$ is finite-dimensional and

$$\dim W \leq \dim V_1 + \cdots + \dim V_m.$$

Then we have

$$\dim(W + V_{m+1}) = \dim W + \dim V_{m+1} - \dim(W \cap V_{m+1}).$$

Since $W \cap V_{m+1}$ is a subspace of V, we have $\dim(W \cap V_{m+1}) \geq 0$. Thus, we have

$$\dim(W + V_{m+1}) \leq \dim W + \dim V_{m+1} \leq \dim V_1 + \cdots + \dim V_m + \dim V_{m+1}.$$

Therefore, $V_1 + \cdots + V_{m+1}$ is finite-dimensional and

$$\dim(V_1 + \cdots + V_{m+1}) \leq \dim V_1 + \cdots + \dim V_m + \dim V_{m+1}. \qquad \square$$

Solution 2C.18 Let $v_1, \ldots, v_n$ be a basis of V. For each $i = 1, \ldots, n$, let $V_i = \operatorname{span} v_i$. Then each V_i is a one-dimensional subspace of V, since it is spanned by a single nonzero vector. Moreover, we have

$$V = V_1 + \cdots + V_n,$$

and the sum is direct because the basis vectors are linearly independent. Therefore, we have

$$V = V_1 \oplus \cdots \oplus V_n. \qquad \square$$

Solution 2C.19 One may guess that the given formula above is true by analogy with the formula for the number of elements in the union of three finite sets. However, this erroneously conflates the idea of subspace sums with the idea of set unions. In fact, the formula is not true in general. For example, let $V = \mathbb{R}^2$, and let

$$V_1 = \operatorname{span}((1, 0)), \quad V_2 = \operatorname{span}((0, 1)), \quad V_3 = \operatorname{span}((1, 1)).$$

It is easy to see that $\dim V_1 = \dim V_2 = \dim V_3 = 1$, but $\dim(V_1 \cap V_2) = \dim(V_1 \cap V_3) = \dim(V_2 \cap V_3) = 0$ and $\dim(V_1 \cap V_2 \cap V_3) = 0$. Thus, the right-hand side of the formula is

$$1 + 1 + 1 - 0 - 0 - 0 + 0 = 3,$$

but the left-hand side is $\dim(V_1 + V_2 + V_3) = \dim \mathbb{R}^2 = 2$. Therefore, the formula is not true in general. $\square$

Solution 2C.20 Note that we need to account for all pairs of subspaces and all pairs of sums of subspaces, since the sum of subspaces is not equivalent to the union of sets. We can use the dimension formula repeatedly to obtain the answer. To that end, define the following subspaces:

$$W_1 = V_1 + V_2, \quad W_2 = V_1 + V_3, \quad W_3 = V_2 + V_3.$$

Applying the dimension formula to each of W_1, W_2, W_3, we have

$$\begin{aligned}
\dim W_1 &= \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2), \\
\dim W_2 &= \dim V_1 + \dim V_3 - \dim(V_1 \cap V_3), \\
\dim W_3 &= \dim V_2 + \dim V_3 - \dim(V_2 \cap V_3).
\end{aligned}$$

Next, we can apply the dimension formula to each of the sums $W_1 + V_3$, $W_2 + V_2$, $W_3 + V_1$:

$$\begin{aligned}
\dim(W_1 + V_3) &= \dim W_1 + \dim V_3 - \dim(W_1 \cap V_3), \\
&= \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2) + \dim V_3 - \dim(W_1 \cap V_3) \\
\dim(W_2 + V_2) &= \dim W_2 + \dim V_2 - \dim(W_2 \cap V_2), \\
&= \dim V_1 + \dim V_3 - \dim(V_1 \cap V_3) + \dim V_2 - \dim(W_2 \cap V_2) \\
\dim(W_3 + V_1) &= \dim W_3 + \dim V_1 - \dim(W_3 \cap V_1), \\
&= \dim V_2 + \dim V_3 - \dim(V_2 \cap V_3) + \dim V_1 - \dim(W_3 \cap V_1).
\end{aligned}$$

Note that $W_1 + V_3 = W_2 + V_2 = W_3 + V_1 = V_1 + V_2 + V_3$. Thus, we can add the three equations above and divide by 3 to obtain the desired formula:

$$\begin{aligned}
\dim(V_1 + V_2 + V_3) = \dim V_1 + \dim V_2 + \dim V_3 \\
- \frac{\dim(V_1 \cap V_2) + \dim(V_1 \cap V_3) + \dim(V_2 \cap V_3)}{3} \\
- \frac{\dim((V_1 + V_2) \cap V_3) + \dim((V_1 + V_3) \cap V_2) + \dim((V_2 + V_3) \cap V_1)}{3}.
\end{aligned}$$

Chapter 3

§3A

Solution 3A.1 Suppose T is linear. Then

$$T(0, 0, 0) = (b, 0),$$

which implies that $b = 0$. Now consider the vector $(1, 1, 1) \in \mathbb{R}^3$. Then

$$T(2, 2, 2) = 2T(1, 1, 1) \implies (2, 12 + 8c) = (2, 12 + 2c).$$

Then $12 + 8c = 12 + 2c$, or $c = 0$.
Conversely, if $b = c = 0$, then $T(x, y, z) = (2x-4y+3z, 6x)$. Let $(x_1, y_1, z_1), (x_2, y_2, z_2) \in \mathbb{R}^3$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned}
T(x_1 + x_2, y_1 + y_2, z_1 + z_2) &= (2(x_1 + x_2) - 4(y_1 + y_2) + 3(z_1 + z_2), 6(x_1 + x_2)) \\
&= (2x_1 + 2x_2 - 4y_1 - 4y_2 + 3z_1 + 3z_2, 6x_1 + 6x_2) \\
&= (2x_1 - 4y_1 + 3z_1, 6x_1) + (2x_2 - 4y_2 + 3z_2, 6x_2) \\
&= T(x_1, y_1, z_1) + T(x_2, y_2, z_2),
\end{aligned}$$

and

$$\begin{aligned}
T(\alpha x_1, \alpha y_1, \alpha z_1) &= (2(\alpha x_1) - 4(\alpha y_1) + 3(\alpha z_1), 6(\alpha x_1)) \\
&= (\alpha(2x_1 - 4y_1 + 3z_1), \alpha(6x_1)) \\
&= \alpha(2x_1 - 4y_1 + 3z_1, 6x_1) \\
&= \alpha T(x_1, y_1, z_1).
\end{aligned}$$

Therefore, T is linear if and only if $b = c = 0$. □

Solution 3A.2 Suppose T is linear. Then for any $p \in \mathbb{R}[X]$, we have $T(2p) = 2Tp$. Looking at the first component of $T(2p)$, we have

$$3(2p)(4) + 5(2p)'(6) + b(2p)(1)(2p)(2) = 6p(4) + 10p'(6) + 4bp(1)p(2).$$

On the other hand, the first component of $2Tp$ is

$$2(3p(4) + 5p'(6) + bp(1)p(2)) = 6p(4) + 10p'(6) + 2bp(1)p(2).$$

Comparing the two expressions, we get $4bp(1)p(2) = 2bp(1)p(2)$ for all $p \in \mathbb{R}[X]$. Since there exist polynomials p such that $p(1)p(2) \neq 0$, we must have $b = 0$.
The second component of $T(2p)$ must be equal to the second component of $2Tp$. We have

$$\int_{-1}^{2} x^3(2p)(x)\,dx + c\sin(2p(0)) = 2\int_{-1}^{2} x^3 p(x)\,dx + 2c\sin(p(0)).$$

The integral terms may be subtracted, so we get $c\sin(2p(0)) = 2c\sin(p(0))$ for all $p \in \mathbb{R}[X]$. Consider $p(x) = \pi/2$. Then $\sin(2p(0)) = \sin(\pi) = 0$ and $2\sin(p(0)) = 2\sin(\pi/2) = 2$. Therefore, $c = 0$.

Conversely, if $b = c = 0$, let $p, q \in \mathbb{R}[X]$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} T(p+q) &= \left(3(p+q)(4) + 5(p+q)'(6), \int_{-1}^{2} x^3(p+q)(x)\,dx\right) \\ &= \left(3p(4) + 3q(4) + 5p'(6) + 5q'(6), \int_{-1}^{2} x^3 p(x)\,dx + \int_{-1}^{2} x^3 q(x)\,dx\right) \\ &= \left(3p(4) + 5p'(6), \int_{-1}^{2} x^3 p(x)\,dx\right) + \left(3q(4) + 5q'(6), \int_{-1}^{2} x^3 q(x)\,dx\right) \\ &= Tp + Tq, \end{aligned}$$

and

$$\begin{aligned} T(\alpha p) &= \left(3(\alpha p)(4) + 5(\alpha p)'(6), \int_{-1}^{2} x^3(\alpha p)(x)\,dx\right) \\ &= \left(\alpha(3p(4) + 5p'(6)), \alpha\int_{-1}^{2} x^3 p(x)\,dx\right) \\ &= \alpha\left(3p(4) + 5p'(6), \int_{-1}^{2} x^3 p(x)\,dx\right) \\ &= \alpha Tp. \end{aligned}$$

Therefore, T is linear if and only if $b = c = 0$. □

Solution 3A.3 Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. Then for each $k = 1, \ldots, n$, we have

$$T(e_k) = (A_{1,k}, A_{2,k}, \ldots, A_{m,k})$$

for some scalars $A_{j,k} \in \mathbb{F}$. Now let $(x_1, \ldots, x_n) \in \mathbb{F}^n$. Then

$$\begin{aligned} T(x_1, \ldots, x_n) &= T(x_1 e_1 + \cdots + x_n e_n) \\ &= x_1 T(e_1) + \cdots + x_n T(e_n) \\ &= x_1 (A_{1,1}, A_{2,1}, \ldots, A_{m,1}) + \cdots + x_n (A_{1,n}, A_{2,n}, \ldots, A_{m,n}) \\ &= (A_{1,1} x_1 + \cdots + A_{1,n} x_n, A_{2,1} x_1 + \cdots + A_{2,n} x_n, \ldots, A_{m,1} x_1 + \cdots + A_{m,n} x_n). \end{aligned}$$

Therefore,

$$T(x_1, \ldots, x_n) = (A_{1,1} x_1 + \cdots + A_{1,n} x_n, A_{2,1} x_1 + \cdots + A_{2,n} x_n, \ldots, A_{m,1} x_1 + \cdots + A_{m,n} x_n).$$

□

Solution 3A.4 Suppose that $a_1 v_1 + \cdots + a_m v_m = 0$ for some scalars $a_1, \ldots, a_m \in \mathbb{F}$. Applying T to both sides, we get

$$a_1 T v_1 + \cdots + a_m T v_m = T(0) = 0.$$

Since $T v_1, \ldots, T v_m$ is linearly independent, it follows that $a_1 = \cdots = a_m = 0$. Therefore, $v_1, \ldots, v_m$ is linearly independent. □

Solution 3A.5 We must first ensure that the operations of addition and scalar multiplication are well-defined, i.e., if $S, T \in \text{Hom}(V, W)$ and $\alpha \in \mathbb{F}$, then $S+T \in \text{Hom}(V, W)$ and $\alpha T \in \text{Hom}(V, W)$. We first show that $S + T$ is linear. Let $v, u \in V$. Then

$$\begin{aligned} (S + T)(v + u) &= S(v + u) + T(v + u) \\ &= (Sv + Su) + (Tv + Tu) \\ &= (Sv + Tv) + (Su + Tu) \\ &= (S + T)v + (S + T)u. \end{aligned}$$

Moreover, for any $v \in V$ and $\alpha \in \mathbb{F}$,

$$\begin{aligned} (S + T)(\alpha v) &= S(\alpha v) + T(\alpha v) \\ &= \alpha S v + \alpha T v \\ &= \alpha (S v + T v) \\ &= \alpha (S + T)v. \end{aligned}$$

This shows that $S + T$ is linear. Now we show that αT is linear. Let $v, u \in V$. Then

$$\begin{aligned} (\alpha T)(v + u) &= \alpha T(v + u) \\ &= \alpha (T v + T u) \\ &= \alpha T v + \alpha T u \\ &= (\alpha T)v + (\alpha T)u. \end{aligned}$$

Also, for any $v \in V$ and $\beta \in \mathbb{F}$,

$$\begin{aligned}(\alpha T)(\beta v) &= \alpha T(\beta v)\\ &= \alpha(\beta T v)\\ &= (\alpha\beta)Tv\\ &= (\beta\alpha)Tv\\ &= \beta(\alpha T v)\\ &= \beta(\alpha T)v.\end{aligned}$$

Therefore, $S + T \in \text{Hom}(V, W)$ and $\alpha T \in \text{Hom}(V, W)$. We now check the vector space axioms:

1. **Commutativity**: Let $S, T \in \text{Hom}(V, W)$. Then for any $v \in V$,

$$\begin{aligned}(S + T)v &= Sv + Tv\\ &= Tv + Sv\\ &= (T + S)v.\end{aligned}$$

2. **Associativity**: Let $R, S, T \in \text{Hom}(V, W)$. Then for any $v \in V$,

$$\begin{aligned}((R + S) + T)v &= (R + S)v + Tv\\ &= (Rv + Sv) + Tv\\ &= Rv + (Sv + Tv)\\ &= Rv + (S + T)v\\ &= (R + (S + T))v.\end{aligned}$$

   Similarly, for any $\alpha, \beta \in \mathbb{F}$ and $T \in \text{Hom}(V, W)$,

$$\begin{aligned}((\alpha\beta)T)v &= (\alpha\beta)(Tv)\\ &= \alpha(\beta T v)\\ &= \alpha((\beta T)v)\\ &= (\alpha(\beta T))v.\end{aligned}$$

3. **Additive identity**: Let $0 \in \text{Hom}(V, W)$ be the zero map defined by $0v = 0$ for all $v \in V$. Then for any $T \in \text{Hom}(V, W)$ and $v \in V$,

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

4. **Inverse**: Let $T \in \text{Hom}(V, W)$. Define $-T \in \text{Hom}(V, W)$ by $(-T)v = -Tv$ for all $v \in V$. Then for any $v \in V$,

$$(T + (-T))v = Tv + (-T)v = Tv - Tv = 0 = 0v.$$

5. **Multiplicative identity**: Let $1 \in \mathbb{F}$ be the multiplicative identity. Then for any $T \in \text{Hom}(V, W)$ and $v \in V$,

$$(1T)v = 1(Tv) = Tv.$$

6. **Distributive properties:** Let $T, S \in \text{Hom}(V, W)$ and $\alpha, \beta \in \mathbb{F}$. Then for any $v \in V$,

$$\begin{aligned}(\alpha(T+S))v &= \alpha((T+S)v)\\ &= \alpha(Tv + Sv)\\ &= \alpha Tv + \alpha Sv\\ &= (\alpha T)v + (\alpha S)v\\ &= ((\alpha T) + (\alpha S))v,\end{aligned}$$

and

$$\begin{aligned}((\alpha+\beta)T)v &= (\alpha+\beta)(Tv)\\ &= \alpha Tv + \beta Tv\\ &= (\alpha T)v + (\beta T)v\\ &= ((\alpha T) + (\beta T))v.\end{aligned}$$

□

Solution 3A.6

(a) Let $T_1 \in \text{Hom}(V, W)$, $T_2 \in \text{Hom}(U, V)$, and $T_3 \in \text{Hom}(X, U)$. Then for any $x \in X$,

$$\begin{aligned}((T_1T_2)T_3)x &= (T_1T_2)(T_3x)\\ &= T_1(T_2(T_3x))\\ &= T_1((T_2T_3)x)\\ &= (T_1(T_2T_3))x.\end{aligned}$$

(b) Let $T \in \text{Hom}(V, W)$. Then for any $v \in V$,

$$(TI)v = T(Iv) = Tv \quad \text{and} \quad (IT)v = I(Tv) = Tv.$$

(c) Let $T, T_1, T_2 \in \text{Hom}(U, V)$ and $S, S_1, S_2 \in \text{Hom}(V, W)$. Then for any $u \in U$,

$$\begin{aligned}((S_1+S_2)T)u &= (S_1+S_2)(Tu)\\ &= S_1(Tu) + S_2(Tu)\\ &= (S_1T)u + (S_2T)u\\ &= (S_1T + S_2T)u,\end{aligned}$$

and

$$\begin{aligned}(S(T_1+T_2))u &= S((T_1+T_2)u)\\ &= S(T_1u + T_2u)\\ &= S(T_1u) + S(T_2u)\\ &= (ST_1)u + (ST_2)u\\ &= (ST_1 + ST_2)u.\end{aligned}$$

□

Solution 3A.7 Suppose $\dim V = 1$ and let $v \in V$ be a nonzero vector. Then $\{v\}$ is a basis of V. Let $T \in \mathrm{End}(V)$. Since $Tv \in V$, there exists $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$. Now let $u \in V$. Then there exists $\alpha \in \mathbb{F}$ such that $u = \alpha v$. Therefore,

$$Tu = T(\alpha v) = \alpha T v = \alpha(\lambda v) = \lambda(\alpha v) = \lambda u. \qquad \square$$

Solution 3A.8

Manual Remark. *This was a particularly difficult exercise because all obvious examples of functions that are homogeneous are also additive. The key idea is to utilize composition of functions.*

Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$\phi(x, y) = \sqrt[3]{x^3 + y^3}.$$

Then for any $a \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^2$,

$$\phi(a(x, y)) = \phi(ax, ay) = \sqrt[3]{(ax)^3 + (ay)^3} = \sqrt[3]{a^3(x^3 + y^3)} = a\sqrt[3]{x^3 + y^3} = a\phi(x, y).$$

However, ϕ is not additive. For example, let $v = (1, 0)$ and $w = (0, 1)$. Then

$$\phi(v + w) = \phi(1, 1) = \sqrt[3]{1^3 + 1^3} = \sqrt[3]{2},$$

but

$$\phi(v) + \phi(w) = \phi(1, 0) + \phi(0, 1) = \sqrt[3]{1^3 + 0^3} + \sqrt[3]{0^3 + 1^3} = 1 + 1 = 2.$$

Therefore, ϕ is not linear. $\square$

Solution 3A.9 Let $\phi : \mathbb{C} \to \mathbb{C}$ be defined by

$$\phi(x + iy) = x.$$

Then for any $w = x_1 + iy_1$ and $z = x_2 + iy_2$ in $\mathbb{C}$,

$$\phi(w + z) = \phi((x_1 + x_2) + i(y_1 + y_2)) = x_1 + x_2 = \phi(w) + \phi(z).$$

However, ϕ is not homogeneous. For example, let $a = i$ and $z = 1$. Then

$$\phi(iz) = \phi(i) = 0,$$

but

$$i\phi(z) = i\phi(1) = i(1) = i.$$

Therefore, ϕ is not linear. $\square$

Solution 3A.10 This is false. Let $q(x) = 1$ and define $T : \mathbb{R}[X] \to \mathbb{R}[X]$ by $Tp = q \circ p$. Then for any $p_1, p_2 \in \mathbb{R}[X]$,

$$T(p_1 + p_2) = q \circ (p_1 + p_2) = 1 \neq 1 + 1 = Tp_1 + Tp_2.$$

Therefore, T is not linear. $\square$

Solution 3A.11 Suppose T is a scalar multiple of the identity, i.e., $T = \lambda I$ for some $\lambda \in \mathbb{F}$. Then for any $S \in \text{End}(V)$,

$$ST = S(\lambda I) = \lambda S = (\lambda I)S = TS.$$

Conversely, suppose $ST = TS$ for every $S \in \text{End}(V)$. We claim that v and Tv are linearly dependent for every $v \in V$. Suppose not. Then there exists $v \in V$ such that v and Tv are linearly independent. Extend $\{v, Tv\}$ to a basis of V. Define $S \in \text{End}(V)$ by sending v to itself and sending $Tv, v_3, \ldots, v_n$ to 0. Then

$$STv = S(Tv) = 0, \quad TSv = Tv \neq 0,$$

which contradicts the assumption that $ST = TS$. Therefore, v and Tv are linearly dependent for every $v \in V$.

Now let $v \in V$ be nonzero. Then there exists $\lambda_v \in \mathbb{F}$ such that $Tv = \lambda_v v$. We claim that λ_v is independent of the choice of v. Let $v_1, \ldots, v_n$ be a basis of V.

- If $n = 1$, then the claim is immediate, since $Tv_1 = \lambda_{v_1} v_1$.
- If $n \geq 2$, let v_i, v_j be two distinct basis vectors. Then their linear independence implies

$$\lambda_{v_i+v_j} v_i + \lambda_{v_i+v_j} v_j = T(v_i + v_j) = Tv_i + Tv_j = \lambda_{v_i} v_i + \lambda_{v_j} v_j.$$

  Since v_i and v_j are linearly independent, it follows that $\lambda_{v_i+v_j} = \lambda_{v_i} = \lambda_{v_j}$. Therefore, λ_v is independent of the choice of v, and we may denote it by λ.

Let $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ be any vector in V. Then

$$Tv = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 Tv_1 + \cdots + \alpha_n Tv_n = \alpha_1 \lambda v_1 + \cdots + \alpha_n \lambda v_n = \lambda(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \lambda v.$$

Therefore, T is a scalar multiple of the identity. □

Solution 3A.12 Let $u \in U$ be such that $Su \neq 0$. Since $U \neq V$, there exists $v \in V$ such that $v \notin U$. Because $v = (v + u) - u \notin U$, then $v + u \notin U$ as well. Therefore,

$$T(u + v) = 0, \quad \text{but} \quad Tu + Tv = Su + 0 = Su \neq 0.$$

Hence, T is not linear. □

Solution 3A.13 Let $u_1, \ldots, u_m$ be a basis of U. Since U is a subspace of V, we can extend this basis to a basis of V, say $u_1, \ldots, u_m, v_1, \ldots, v_n$. Define $T : V \to W$ by

$$Tu_i = Su_i \text{ for all } i = 1, \ldots, m, \quad \text{and} \quad Tv_k = 0 \text{ for all } k = 1, \ldots, n.$$

By the linear map lemma, $T \in \text{Hom}(V, W)$. Moreover, for any $u \in U$, there exist scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$ such that $u = \sum_{i=1}^{m} \alpha_i u_i$. Then

$$Tu = \sum_{i=1}^{m} \alpha_i Tu_i = \sum_{i=1}^{m} \alpha_i Su_i = Su$$

for all $u \in U$. □

Solution 3A.14 Let $v_1, \dots, v_n$ be a basis of V. By Exercise 2A17, there exist a sequence $w_1, w_2, \dots$ of vectors in W such that $w_1, \dots, w_m$ is linearly independent for each $m \in \mathbb{N}$. For each $i \in \mathbb{N}$, define $T_i \in \text{Hom}(V, W)$ by

$$T_i v_j = \begin{cases} w_i & \text{if } j = 1, \\ 0 & \text{if } j \neq 1. \end{cases}$$

In other words, T_i sends the first basis vector v_1 to w_i and all other basis vectors to 0. We now have a sequence of linear maps $T_1, T_2, \dots$ in $\text{Hom}(V, W)$. Suppose that

$$\sum_{\ell=1}^{m} \alpha_\ell T_\ell = 0$$

for some scalars $\alpha_1, \dots, \alpha_m \in \mathbb{F}$. Applying both sides to v_1, we get

$$\sum_{\ell=1}^{m} \alpha_\ell T_\ell v_1 = \sum_{\ell=1}^{m} \alpha_\ell w_\ell = 0.$$

Since $w_1, \dots, w_m$ is linearly independent, it follows that $\alpha_1 = \cdots = \alpha_m = 0$. It follows that $T_1, \dots, T_m$ is linearly independent in $\text{Hom}(V, W)$ for each $m \in \mathbb{N}$. Because there exists a sequence $T_1, T_2, \dots$ of linearly independent linear maps in $\text{Hom}(V, W)$ where $T_1, \dots, T_m$ is linearly independent for each $m \in \mathbb{N}$, it follows by Exercise 2A17 that $\text{Hom}(V, W)$ is infinite-dimensional. □

Solution 3A.15 Since $v_1, \dots, v_m$ is linearly dependent, there exist scalars $a_1, \dots, a_m \in \mathbb{F}$, not all zero, such that

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

Fix some nonzero a_i among $a_1, \dots, a_m$. Let $w_1, \dots, w_m \in W$ be defined by

$$w_k = \begin{cases} 0 & \text{if } k \neq i, \\ w & \text{if } k = i, \end{cases}$$

where w is any nonzero vector in W. Suppose for contradiction that there exists $T \in \text{Hom}(V, W)$ such that $T v_k = w_k$ for each $k = 1, \dots, m$. Then

$$T(a_1 v_1 + \cdots + a_m v_m) = a_1 T v_1 + \cdots + a_m T v_m = a_1 w_1 + \cdots + a_m w_m.$$

Since $a_1 v_1 + \cdots + a_m v_m = 0$, we have $T(0) = 0$, so

$$0 = a_1 w_1 + \cdots + a_m w_m.$$

However, by definition of w_k, we have $a_k w_k = a_i w$ if $k = i$ and $a_k w_k = 0$ if $k \neq i$. Therefore, the sum $a_1 w_1 + \cdots + a_m w_m$ is a nonzero scalar multiple of w, which is nonzero since $w \neq 0$. This is a contradiction. Therefore, no such $T \in \text{Hom}(V, W)$ exists. □

Solution 3A.16 Let v_1, v_2 be linearly independent vectors in V. We may extend this list to a basis of V, say $v_1, v_2, \ldots, v_n$. Define $S, T \in \text{End}(V)$ by

$$Sv_1 = v_2, \quad Sv_2 = v_1, \quad Sv_k = 0 \text{ for } k > 2,$$

and

$$Tv_1 = v_1, \quad Tv_2 = 0, \quad Tv_k = 0 \text{ for } k > 2.$$

Then

$$STv_1 = S(Tv_1) = Sv_1 = v_2,$$

but

$$TSv_1 = T(Sv_1) = Tv_2 = 0.$$

Therefore, $ST \neq TS$. □

Solution 3A.17

Axler Note. *A subspace $\mathscr{E}$ of* End(V) *is called a two-sided ideal of* End(V) *if $TE \in \mathscr{E}$ and $ET \in \mathscr{E}$ for all $E \in \mathscr{E}$ and all $T \in$* End(V).

Let $\mathscr{E}$ be a two-sided ideal of End(V). If $\mathscr{E} = \{0\}$, then we are done. Otherwise, there exists $E \in \mathscr{E}$ such that $E \neq 0$. Then there exists $v_1 \in V$ such that $Ev_1 \neq 0$. Since V is finite-dimensional, we may extend the list $\{v\}_1$ to a basis of V, say $v_1, v_2, \ldots, v_n$. Let $w_1 = Ev_1 \neq 0$. We may also extend the list $\{w\}_1$ to a basis of V, say $w_1, w_2, \ldots, w_n$. Define $T_1 \in \text{End}(V)$ by

$$T_1 w_1 = v_1, \quad T_1 w_k = 0 \text{ for } k > 1.$$

Then $T_1 E \in \mathscr{E}$ since $\mathscr{E}$ is a two-sided ideal. Observe that $T_1 E v_1 = T_1(Ev_1) = T_1 w_1 = v_1 \neq 0$. Therefore, $T_1 E \neq 0$. Consider $S_1 \in \text{End}(V)$ defined by

$$S_1 v_1 = v_1, \quad S_1 v_k = 0 \text{ for } k > 1.$$

Then $T_1 E S_1 \in \mathscr{E}$ since $\mathscr{E}$ is a two-sided ideal. Observe that $T_1 E S_1 v_1 = T_1 E v_1 = T_1 w_1 = v_1 \neq 0$. Therefore, $T_1 E S_1 \neq 0$. We now discuss the construction of the identity map $I \in \text{End}(V)$.

Construct $T_i, S_i \in \text{End}(V)$ for each $i = 1, \ldots, n$ as follows:

$$T_i w_1 = v_i, \quad T_i w_k = 0 \text{ for } k \neq 1,$$

and

$$S_i v_i = v_1, \quad S_i v_k = 0 \text{ for } k \neq i.$$

Then $T_i E S_i \in \mathscr{E}$ and $T_i E S_i v_i = v_i \neq 0$. Since $\mathscr{E}$ is a subspace, it contains all linear combinations of these maps. In particular, for any $v \in V$, we can write $v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ for some scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}$. Observe that

$$T_i E S_i v = T_i E S_i(\alpha_1 v_1) + \cdots + T_i E S_i(\alpha_i v_i) + \cdots + T_i E S_i(\alpha_n v_n) = \alpha_i v_i.$$

Consider the map

$$E^* = T_1 E S_1 + T_2 E S_2 + \cdots + T_n E S_n \in \mathscr{E},$$

Then for any $v = \alpha_1 v_1 + \cdots + \alpha_n v_n \in V$,

$$E^* v = T_1 E S_1 v + T_2 E S_2 v + \cdots + T_n E S_n v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = v.$$

Hence, $E^* = I \in \mathscr{E}$. Since $\mathscr{E}$ is a two-sided ideal, $I = E^* \in \mathscr{E}$ implies that for any $T \in \text{End}(V)$, we have $TI = T \in \mathscr{E}$. Therefore, $\mathscr{E} = \text{End}(V)$. □

§3B

Solution 3B.1 Let $T \in \text{Hom}(\mathbb{R}^5, \mathbb{R}^2)$ be defined by

$$T(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2).$$

Then $\ker T = \{(0, 0, x_3, x_4, x_5) : x_3, x_4, x_5 \in \mathbb{R}\}$, which has dimension 3, and $\text{im}\, T = \mathbb{R}^2$, which has dimension 2. □

Solution 3B.2 Let $v \in V$. Then

$$(ST)^2 v = ST(STv) = S(T(STv)).$$

Since $STv \in \text{im}\, S$ and $\text{im}\, S \subseteq \ker T$, we have $T(STv) = 0$. Therefore,

$$(ST)^2 v = S(0) = 0.$$

Since v was arbitrary, it follows that $(ST)^2 = 0$. □

Solution 3B.3

(a) The list $v_1, \ldots, v_m$ spans V if and only if $\text{im}\, T = V$, i.e., T is surjective.

(b) The list $v_1, \ldots, v_m$ is linearly independent if and only if $\ker T = \{0\}$, i.e., T is injective.

□

Solution 3B.4 Let $T_1, T_2 \in \text{Hom}(\mathbb{R}^5, \mathbb{R}^4)$ be defined by

$$T_1(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2, 0, 0)$$

and

$$T_2(x_1, x_2, x_3, x_4, x_5) = (0, 0, x_3, x_4).$$

Then $\ker T_1 = \{(0, 0, x_3, x_4, x_5) : x_3, x_4, x_5 \in \mathbb{R}\}$ and $\ker T_2 = \{(x_1, x_2, 0, 0, x_5) : x_1, x_2, x_5 \in \mathbb{R}\}$ both have dimension 3, so T_1 and T_2 are in the set $\{T \in \text{Hom}(\mathbb{R}^5, \mathbb{R}^4) : \dim \ker T > 2\}$. However,

$$(T_1 + T_2)(x_1, x_2, x_3, x_4, x_5) = (x_1, x_2, x_3, x_4),$$

which has null space $\ker(T_1 + T_2) = \{(0, 0, 0, 0, x_5) : x_5 \in \mathbb{R}\}$ of dimension 1. Therefore,

$$\dim \ker(T_1 + T_2) = 1 < 2,$$

so $T_1 + T_2$ is not in the set. Hence, the set is not closed under addition and is not a subspace of $\text{Hom}(\mathbb{R}^5, \mathbb{R}^4)$. □

Solution 3B.5 Let v_1, v_2, v_3, v_4 be the standard basis of $\mathbb{R}^4$. Define $T \in \text{End}(\mathbb{R}^4)$ by

$$Tv_1 = v_3, \quad Tv_2 = v_4, \quad Tv_3 = 0, \quad Tv_4 = 0.$$

Then for any $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4$,

$$Tx = x_1Tv_1 + x_2Tv_2 + x_3Tv_3 + x_4Tv_4 = x_1v_3 + x_2v_4.$$

Then $\text{im}\, T = \text{span}(v_3, v_4)$. If $Tx = 0$, then $x_1v_3 + x_2v_4 = 0$, which implies $x_1 = x_2 = 0$. Therefore, $\ker T = \{(0, 0, x_3, x_4) : x_3, x_4 \in \mathbb{R}\} = \text{span}(v_3, v_4)$. Hence, $\text{im}\, T = \ker T$. □

Solution 3B.6 Suppose for contradiction that there exists $T \in \text{End}(\mathbb{R}^5)$ such that $\text{im}\, T = \ker T$. Then by the fundamental theorem of linear maps,

$$\dim \mathbb{R}^5 = \dim \ker T + \dim \text{im}\, T = 2 \dim \text{im}\, T.$$

Therefore, $5 = 2 \dim \text{im}\, T$, which is impossible since 5 is odd and $2 \dim \text{im}\, T$ is even. Hence, no such T exists. □

Solution 3B.7 Let $\dim V = n$ and $\dim W = m$ with $2 \leq n \leq m$. Let $v_1, v_2, \ldots, v_n$ be a basis of V, and let $w_1, w_2, \ldots, w_m$ be a basis of W. Define $T_1, T_2 \in \text{Hom}(V, W)$ by

$$T_1v_1 = w_1, \quad T_1v_k = 0 \text{ for } k \geq 2,$$

and

$$T_2v_1 = 0, \quad T_2v_k = w_k \text{ for } k \geq 2.$$

Then $v_2 \in \ker T_1$ and $v_1 \in \ker T_2$, so both T_1 and T_2 are not injective. However,

$$(T_1 + T_2)v_1 = T_1v_1 + T_2v_1 = w_1 + 0 = w_1 \neq 0,$$

and for $k \geq 2$,

$$(T_1 + T_2)v_k = T_1v_k + T_2v_k = 0 + w_k = w_k \neq 0.$$

Suppose $v \in V$ is such that $(T_1 + T_2)v = 0$. Then v can be expressed as a linear combination of the basis vectors:

$$v = \alpha_1v_1 + \alpha_2v_2 + \cdots + \alpha_nv_n.$$

Then

$$(T_1 + T_2)v = \alpha_1w_1 + \alpha_2w_2 + \cdots + \alpha_nw_n = 0.$$

Since $w_1, w_2, \ldots, w_m$ are linearly independent, it follows that $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$, hence $v = 0$. Therefore, $T_1 + T_2$ is injective. Since T_1 and T_2 are not injective, but their sum is injective, the set $\{T \in \text{Hom}(V, W) : T \text{ is not injective}\}$ is not closed under addition and thus is not a subspace of $\text{Hom}(V, W)$. □

Solution 3B.8 We proceed similarly to the previous exercise. Let $\dim V = n$ and $\dim W = m$ with $n \geq m \geq 2$. Let $v_1, v_2, \ldots, v_n$ be a basis of V, and let $w_1, w_2, \ldots, w_m$ be a basis of W. Define $T_1, T_2 \in \text{Hom}(V, W)$ by

$$T_1 v_1 = w_1, \quad T_1 v_k = 0 \text{ for } k \geq 2,$$

and

$$T_2 v_1 = 0, \quad T_2 v_k = w_k \text{ for } k = 2, \ldots, m, \quad T_2 v_k = 0 \text{ for } k > m.$$

Then $\text{im}\, T_1 = \text{span}\, w_1$ and $\text{im}\, T_2 = \text{span}\, w_2, \ldots, w_m$, both of which are proper subspaces of W, so T_1 and T_2 are not surjective. However,

$$(T_1 + T_2)v_1 = T_1 v_1 + T_2 v_1 = w_1 + 0 = w_1,$$

and for $k = 2, \ldots, m$,

$$(T_1 + T_2)v_k = T_1 v_k + T_2 v_k = 0 + w_k = w_k.$$

Therefore, $\text{im}(T_1 + T_2) = \text{span}\, w_1, w_2, \ldots, w_m = W$, so $T_1 + T_2$ is surjective. Since T_1 and T_2 are not surjective, but their sum is surjective, the set $\{T \in \text{Hom}(V, W) : T \text{ is not surjective}\}$ is not closed under addition and thus is not a subspace of $\text{Hom}(V, W)$. □

Solution 3B.9 Suppose $a_1, \ldots, a_n \in \mathbb{F}$ are such that

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

By linearity of T, we have

$$T(a_1 v_1 + \cdots + a_n v_n) = 0.$$

Injectivity of T implies that $\ker T = \{0\}$, so

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

Since $v_1, \ldots, v_n$ is linearly independent, it follows that $a_1 = \cdots = a_n = 0$. Therefore, $T v_1, \ldots, T v_n$ is linearly independent in W. □

Solution 3B.10 Let $w \in \text{im}\, T$. Then there exists $v \in V$ such that $Tv = w$. Since $v_1, \ldots, v_n$ spans V, we can write

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

for some scalars $a_1, \ldots, a_n \in \mathbb{F}$. By linearity of T, we have

$$w = Tv = T(a_1 v_1 + \cdots + a_n v_n) = a_1 T v_1 + \cdots + a_n T v_n.$$

Therefore, w is a linear combination of $T v_1, \ldots, T v_n$. Since w was arbitrary in $\text{im}\, T$, it follows that $T v_1, \ldots, T v_n$ spans $\text{im}\, T$. □

Solution 3B.11

Manual Remark. *Note that this exercise proves the existence of a subspace U of V such that $V = U \oplus \ker T$.*

By 2.33, because $\ker T$ is a subspace of V with V finite dimensional, there exists a subspace U of V such that $V = \ker T \oplus U$. Then $U \cap \ker T = \{0\}$. Additionally, every $v \in V$ can be written uniquely as $v = n + u$ with $n \in \ker T$ and $u \in U$. Hence,

$$\begin{aligned}\operatorname{im} T &= \{Tv : v \in V\} \\ &= \{T(n+u) : n \in \ker T, u \in U\} \\ &= \{Tn + Tu : n \in \ker T, u \in U\} \\ &= \{Tu : u \in U\}.\end{aligned}$$

□

Solution 3B.12 We may write the null space of T as

$$\ker T = \{(5x_2, x_2, 7x_4, x_4) : x_2, x_4 \in \mathbb{F}\}.$$

From this, we can see that $\ker T$ is spanned by the vectors $(5, 1, 0, 0)$ and $(0, 0, 7, 1)$. Therefore, $\dim \ker T = 2$. By the fundamental theorem of linear maps, we have

$$\dim \mathbb{F}^4 = \dim \ker T + \dim \operatorname{im} T \implies 4 = 2 + \dim \operatorname{im} T.$$

Solving for $\dim \operatorname{im} T$, we find that $\dim \operatorname{im} T = 2$. Since $\dim \mathbb{F}^2 = 2$, it follows that $\operatorname{im} T = \mathbb{F}^2$. Therefore, T is surjective. □

Solution 3B.13 By the fundamental theorem of linear maps, we have

$$\dim \mathbb{R}^8 = \dim \ker T + \dim \operatorname{im} T \implies 8 = 3 + \dim \operatorname{im} T.$$

Solving for $\dim \operatorname{im} T$, we find that $\dim \operatorname{im} T = 5$. Since $\dim \mathbb{R}^5 = 5$, it follows that $\operatorname{im} T = \mathbb{R}^5$. Therefore, T is surjective. □

Solution 3B.14 We can express the null space as

$$\ker T = \{(3x_2, x_2, x_4, x_4, x_4) : x_2, x_4 \in \mathbb{F}\}.$$

From this, we can see that $\ker T$ is spanned by the vectors $(3, 1, 0, 0, 0)$ and $(0, 0, 1, 1, 1)$. Therefore, $\dim \ker T = 2$. By the fundamental theorem of linear maps, we have

$$\dim \mathbb{F}^5 = \dim \ker T + \dim \operatorname{im} T \implies 5 = 2 + \dim \operatorname{im} T.$$

Solving for $\dim \operatorname{im} T$, we find that $\dim \operatorname{im} T = 3$. However, $\dim \mathbb{F}^2 = 2$, which is less than 3. This is a contradiction, as the dimension of the range cannot exceed the dimension of the target space. Therefore, no such linear map exists. □

Solution 3B.15

Manual Remark. *The spanning list in the solution is actually a basis of V.*

Let $T \in \text{End}(V)$ be such that $\ker T$ and $\text{im}\, T$ are both finite-dimensional. Let $u_1, \ldots, u_m$ be a basis of $\ker T$, and let $w_1, \ldots, w_n$ be a basis of $\text{im}\, T$. Select $v_j \in V$ be such that $Tv_j = w_j$ for $j = 1, \ldots, n$. We claim that the list $u_1, \ldots, u_m, v_1, \ldots, v_n$ spans V.

Let $v \in V$. Then $Tv \in \text{im}\, T$, so we may write

$$Tv = \alpha_1 w_1 + \cdots + \alpha_n w_n$$

for some scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Consider the vector

$$u = v - (\alpha_1 v_1 + \cdots + \alpha_n v_n).$$

Then

$$Tu = Tv - (\alpha_1 Tv_1 + \cdots + \alpha_n Tv_n) = Tv - (\alpha_1 w_1 + \cdots + \alpha_n w_n) = 0.$$

Therefore, $u \in \ker T$, and we can write

$$u = \beta_1 u_1 + \cdots + \beta_m u_m$$

for some scalars $\beta_1, \ldots, \beta_m \in \mathbb{F}$. Thus,

$$v = u + (\alpha_1 v_1 + \cdots + \alpha_n v_n) = \beta_1 u_1 + \cdots + \beta_m u_m + \alpha_1 v_1 + \cdots + \alpha_n v_n,$$

which shows that the list spans V. Hence, V is finite-dimensional. □

Solution 3B.16 Suppose there exists an injective linear map $T \in \text{Hom}(V, W)$. By contraposition of 3.22, an injective linear map $T \in \text{Hom}(V, W)$ implies that $\dim V \leq \dim W$.

Conversely, suppose $\dim V \leq \dim W$. Let $n = \dim V$ and $m = \dim W$. Choose a basis $\{v_1, \ldots, v_n\}$ of V. Let $\{w_1, \ldots, w_m\}$ be a basis of W. Define a linear map $T \in \text{Hom}(V, W)$ by

$$Tv_i = w_i \quad \text{for } i = 1, \ldots, n.$$

Since $n \leq m$, we can assign each basis vector of V to a distinct basis vector of W. To show that T is injective, let $v = \alpha_1 v_1 + \cdots + \alpha_n v_n \in V$ such that $Tv = 0$. Then

$$Tv = \alpha_1 Tv_1 + \cdots + \alpha_n Tv_n = \alpha_1 w_1 + \cdots + \alpha_n w_n = 0.$$

Since $w_1, \ldots, w_n$ are linearly independent, it follows that $\alpha_1 = \cdots = \alpha_n = 0$. Therefore, $v = 0$, and T is injective. Hence, there exists an injective linear map from V to W if and only if $\dim V \leq \dim W$. □

Solution 3B.17 Suppose there exists a surjective linear map $T \in \text{Hom}(V, W)$. By contraposition of 3.24, a surjective linear map $T \in \text{Hom}(V, W)$ implies that $\dim V \geq \dim W$.

Conversely, suppose $\dim V \geq \dim W$. Let $n = \dim V$ and $m = \dim W$. Choose a basis $\{v_1, \ldots, v_n\}$ of V. Let $\{w_1, \ldots, w_m\}$ be a basis of W. Define a linear map $T \in \mathrm{Hom}(V, W)$ by

$$Tv_i = w_i \quad \text{for } i = 1, \ldots, m, \quad Tv_i = 0 \quad \text{for } i = m+1, \ldots, n.$$

To show that T is surjective, let $w \in W$. Then w can be expressed as a linear combination of the basis vectors of W:

$$w = \beta_1 w_1 + \cdots + \beta_m w_m.$$

Consider the vector $v = \beta_1 v_1 + \cdots + \beta_m v_m \in V$. Then

$$Tv = \beta_1 Tv_1 + \cdots + \beta_m Tv_m = \beta_1 w_1 + \cdots + \beta_m w_m = w.$$

Since w was arbitrary in W, it follows that $\operatorname{im} T = W$. Therefore, T is surjective. Hence, there exists a surjective linear map from V onto W if and only if $\dim V \geq \dim W$. □

Solution 3B.18 Suppose there exists $T \in \mathrm{Hom}(V, W)$ such that $\ker T = U$. By the fundamental theorem of linear maps, we have

$$\dim V = \dim \ker T + \dim \operatorname{im} T = \dim U + \dim \operatorname{im} T.$$

Since $\operatorname{im} T$ is a subspace of W, we have $\dim \operatorname{im} T \leq \dim W$. Therefore,

$$\dim V = \dim U + \dim \operatorname{im} T \leq \dim U + \dim W.$$

Rearranging this inequality gives

$$\dim U \geq \dim V - \dim W.$$

Conversely, suppose $\dim U \geq \dim V - \dim W$. Let $n = \dim V$, $m = \dim W$, and $k = \dim U$. Let $u_1, \ldots, u_k$ be a basis of U. Extend this basis to a basis of V by adding vectors $v_1, \ldots, v_{n-k}$. Define a linear map $T \in \mathrm{Hom}(V, W)$ by

$$Tu_i = 0 \quad \text{for } i = 1, \ldots, k, \quad Tv_j = w_j \quad \text{for } j = 1, \ldots, n-k,$$

where $w_1, \ldots, w_{n-k}$ are linearly independent vectors in W. Since $\dim U = k$ and $\dim V - \dim W = n - m$, we have $k \geq n - m$, which implies that $n - k \leq m$. Therefore, we can choose $w_1, \ldots, w_{n-k}$ to be linearly independent in W. To show that $\ker T = U$, let $v \in \ker T$. Then $Tv = 0$. Express v as a linear combination of the basis vectors of V:

$$v = \alpha_1 u_1 + \cdots + \alpha_k u_k + \beta_1 v_1 + \cdots + \beta_{n-k} v_{n-k}.$$

Applying T to both sides, we have

$$Tv = \alpha_1 Tu_1 + \cdots + \alpha_k Tu_k + \beta_1 Tv_1 + \cdots + \beta_{n-k} Tv_{n-k} = 0 + \beta_1 w_1 + \cdots + \beta_{n-k} w_{n-k} = 0.$$

Since $w_1, \ldots, w_{n-k}$ are linearly independent, it follows that $\beta_1 = \cdots = \beta_{n-k} = 0$. Therefore,

$$v = \alpha_1 u_1 + \cdots + \alpha_k u_k \in U.$$

Hence, $\ker T \subseteq U$. Conversely, if $v \in U$, then $Tv = 0$ by the definition of T. Therefore, $U \subseteq \ker T$. Combining these two inclusions, we have $\ker T = U$. Thus, there exists $T \in \mathrm{Hom}(V, W)$ such that $\ker T = U$ if and only if $\dim U \geq \dim V - \dim W$. □

Solution 3B.19 Suppose T is injective. Since $\operatorname{im} T$ is a subspace of W, it is finite-dimensional. Let $\dim \operatorname{im} T = n$. If there is no linearly independent list of n vectors in V, then V is finite-dimensional, and $\dim V < n$, which contradicts the fundamental theorem of linear maps. Therefore, there exists a linearly independent list $v_1, \ldots, v_n$ of vectors in V.
Since T is injective, the list $Tv_1, \ldots, Tv_n$ is linearly independent in $\operatorname{im} T$ by Exercise 3B9. Additionally, the list has length n and so is a basis of $\operatorname{im} T$. By the linear map lemma, there exists $S_0 \in \operatorname{Hom}(\operatorname{im} T, V)$ such that $S_0(Tv_i) = v_i$ for all $i = 1, \ldots, n$. By Exercise 3A13, there exists an extension $S \in \operatorname{Hom}(W, V)$ of S_0. Since

$$(ST)v_i = S(Tv_i) = S_0(Tv_i) = v_i$$

for all $i = 1, \ldots, n$, ST acts as the identity on a basis of V. By the linear map lemma, this map is unique, hence ST is the identity operator on V.

Conversely, suppose there exists $S \in \operatorname{Hom}(W, V)$ such that ST is the identity operator on V. Let $v \in \ker T$. Then $Tv = 0$, and applying S gives

$$STv = S(0) = 0.$$

Since ST is the identity operator on V, we have $v = STv = 0$. Therefore, $\ker T = \{0\}$, which means T is injective. □

Solution 3B.20 Suppose T is surjective, and let $w_1, \ldots, w_n$ be a basis for W. By surjectivity of T, there are vectors $v_1, \ldots, v_n \in V$ where $Tv_i = w_i$ for all $i = 1, \ldots, n$. By the linear map lemma, there exists $S \in \operatorname{Hom}(W, V)$ such that

$$Sw_i = v_i \quad \text{for } i = 1, \ldots, n.$$

Then

$$TSw_i = T(Sw_i) = Tv_i = w_i \quad \text{for } i = 1, \ldots, n.$$

By the linear map lemma, because TS agrees with the identity on a basis of W, it must be the identity operator on W.
Conversely, suppose there exists $S \in \operatorname{Hom}(W, V)$ such that TS is the identity operator on W. Let $w \in W$. Then

$$TSw = w.$$

This means that for every $w \in W$, there exists a vector $v = Sw \in V$ such that $Tv = T(Sw) = w$. Therefore, $\operatorname{im} T = W$, which means that T is surjective. □

Solution 3B.21 Let $X = \{v \in V : Tv \in U\}$. We first show that X is a subspace of V. Let $v_1, v_2 \in X$ and $\alpha, \beta \in \mathbb{F}$. Then

$$T(\alpha v_1 + \beta v_2) = \alpha Tv_1 + \beta Tv_2.$$

Since $Tv_1, Tv_2 \in U$ and U is a subspace of W, it follows that $\alpha Tv_1 + \beta Tv_2 \in U$. Therefore, $\alpha v_1 + \beta v_2 \in X$, and thus X is a subspace of V.
Because V is finite-dimensional, so are $\ker T$ and X. Let $v_1, \ldots, v_m$ be a basis for $\ker T$. Since U is a subspace of W, $0 \in U$, so $\ker T \subseteq X$. We may then extend $v_1, \ldots, v_m$ to

a basis $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ of X. If $w \in U \cap \operatorname{im} T$, then there exists $v \in V$ such that $Tv = w$, and since $v_1, \ldots, v_n$ is a basis of X, we may write

$$w = Tv = T\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i T v_i = \sum_{i=m+1}^{n} \alpha_i T v_i$$

for scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Then $Tv_{m+1}, \ldots, Tv_n$ span $U \cap \operatorname{im} T$. Additionally, if $\sum_{i=m+1}^{n} \alpha_i T v_i = 0$, then

$$T\left(\sum_{i=m+1}^{n} \alpha_i v_i\right) = \sum_{i=m+1}^{n} \alpha_i T v_i = 0,$$

which means that $\sum_{i=m+1}^{n} \alpha_i v_i \in \ker T$, contradicting the fact that $v_1, \ldots, v_n$ was linearly independent unless all $\alpha_{m+1}, \ldots, \alpha_n$ are zero. Therefore, $Tv_{m+1}, \ldots, Tv_n$ is linearly independent in $U \cap \operatorname{im} T$, so it is a basis for $U \cap \operatorname{im} T$. Hence,

$$\dim(U \cap \operatorname{im} T) = n - m = \dim X - \dim \ker T,$$

where rearranging gives the desired formula. □

Solution 3B.22 Let $u \in U$ such that $u \in \ker ST$. Then $STu = 0$, which implies that $Tu \in \ker S$. Therefore, we can express $\ker ST$ as

$$\ker ST = \{u \in U : Tu \in \ker S\}.$$

Observe that this looks similar to the set in Exercise 3b21, where T is a linear map from U to V and $\ker S$ is a subspace of V. By applying the result from Exercise 3b21, we have

$$\dim \ker ST = \dim \ker T + \dim(\ker S \cap \operatorname{im} T).$$

Since $\ker S \cap \operatorname{im} T$ is a subspace of $\ker S$, we have

$$\dim(\ker S \cap \operatorname{im} T) \leq \dim \ker S.$$

Therefore,

$$\dim \ker ST = \dim \ker T + \dim(\ker S \cap \operatorname{im} T) \leq \dim \ker T + \dim \ker S. \qquad \square$$

Solution 3B.23 Let $u \in U$ such that $STu \in \operatorname{im} ST$. Then $STu = S(Tu)$, which implies that $Tu \in \operatorname{im} T$. Therefore, we can express $\operatorname{im} ST$ as

$$\operatorname{im} ST = \{S(Tu) : u \in U\}.$$

Since $Tu \in \operatorname{im} T$, we can rewrite this as

$$\operatorname{im} ST = \{Sv : v \in \operatorname{im} T\}.$$

This shows that $\operatorname{im} ST$ is a subset of $\operatorname{im} S$, and hence

$$\dim \operatorname{im} ST \leq \dim \operatorname{im} S.$$

Now let $v_1, \ldots, v_m$ be a basis of $\operatorname{im} T$. Then for any $u \in U$, we can express Tu as a linear combination of the basis vectors:

$$Tu = \alpha_1 v_1 + \cdots + \alpha_m v_m$$

for some scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$. Applying S to both sides, we have

$$STu = S(Tu) = S(\alpha_1 v_1 + \cdots + \alpha_m v_m) = \alpha_1 S v_1 + \cdots + \alpha_m S v_m.$$

This shows that $\operatorname{im} ST$ is spanned by the vectors $Sv_1, \ldots, Sv_m$. Therefore,

$$\dim \operatorname{im} ST \leq m = \dim \operatorname{im} T.$$

Combining the two inequalities, we have

$$\dim \operatorname{im} ST \leq \min(\dim \operatorname{im} S, \dim \operatorname{im} T). \qquad \square$$

Solution 3B.24

(a) Since $ST = 0$, then $S(Tv) = 0$ for all $v \in V$. This implies that $\operatorname{im} T \subseteq \ker S$. By the fundamental theorem of linear maps, we have

$$5 = \dim \ker S + \dim \operatorname{im} S \geq \dim \operatorname{im} T + \dim \operatorname{im} S.$$

Now, by Exercise 3B23, we have

$$\dim \operatorname{im} TS \leq \min(\dim \operatorname{im} T, \dim \operatorname{im} S).$$

Suppose that $\dim \operatorname{im} TS > 2$. Then both $\dim \operatorname{im} T$ and $\dim \operatorname{im} S$ must be greater than 2, or equivalently, $\dim \operatorname{im} T \geq 3$ and $\dim \operatorname{im} S \geq 3$. This implies that

$$\dim \operatorname{im} T + \dim \operatorname{im} S \geq 3 + 3 = 6 > 5,$$

which is a contradiction. Therefore, $\dim \operatorname{im} TS \leq 2$.

(b) Let v_1, v_2, v_3, v_4, v_5 be a basis of $\mathbb{F}^5$. Define $S, T \in \operatorname{End}(\mathbb{F}^5)$ by

$$Sv_1 = v_1, \quad Sv_2 = v_2, \quad Sv_3 = Sv_4 = Sv_5 = 0,$$

and

$$Tv_1 = v_3, \quad Tv_2 = v_4, \quad Tv_3 = Tv_4 = Tv_5 = 0.$$

Then $ST = 0$ since $S(Tv_i) = S(0) = 0$ for all $i = 1, \ldots, 5$. Now, we compute $\operatorname{im} TS$. We have

$$TSv_1 = T(Sv_1) = Tv_1 = v_3, \quad TSv_2 = T(Sv_2) = Tv_2 = v_4, \quad TSv_3 = TSv_4 = TSv_5 = 0.$$

Therefore, $\operatorname{im} TS = \operatorname{span}(v_3, v_4)$, which has dimension 2. Thus, we have constructed an example of $S, T \in \operatorname{End}(\mathbb{F}^5)$ with $ST = 0$ and $\dim \operatorname{im} TS = 2$.

$\square$

Solution 3B.25

Manual Remark. *This exercise requires showing that some* $R \in \operatorname{Hom}(\operatorname{im} S, W)$ *is well-defined. An issue arises when* T *is not injective: for some* $w \in \operatorname{im} S$*, there may be distinct* $v_1, v_2 \in V$ *such that* $Sv_1 = Sv_2 = w$*. In this case, we must show that* $Tv_1 = Tv_2$*.*

Suppose $\ker S \subseteq \ker T$. Define a linear map $R : \operatorname{im} S \to W$ by $R(Sv) = Tv$ for all $v \in V$. We need to show that R is well-defined. Suppose $Sv_1 = Sv_2$ for some $v_1, v_2 \in V$. Then $S(v_1 - v_2) = 0$, which implies that $v_1 - v_2 \in \ker S$. Since $\ker S \subseteq \ker T$, we have $v_1 - v_2 \in \ker T$, which means that $T(v_1 - v_2) = 0$. Therefore, $Tv_1 = Tv_2$, and hence $R(Sv_1) = R(Sv_2)$. Thus, R is well-defined.

By Exercise 3A13, we can extend R to some $E \in \operatorname{End}(W)$. Then for any $v \in V$, we have

$$Tv = R(Sv) = E(Sv),$$

and hence $T = ES$.

Conversely, suppose there exists $E \in \operatorname{End}(W)$ such that $T = ES$. Let $v \in \ker S$. Then $Sv = 0$, and applying E gives

$$Tv = E(Sv) = E(0) = 0.$$

Therefore, $v \in \ker T$, and hence $\ker S \subseteq \ker T$. □

Solution 3B.26

Manual Remark. *From the previous exercise, the reader may be tempted to show that a map is well-defined again. However, this is not necessary here, since we are defining a map on the entire space* V*, not just on* $\operatorname{im} S$*.*

Suppose $\operatorname{im} S \subseteq \operatorname{im} T$. Let $v_1, \ldots, v_n$ be a basis of V. For each $i = 1, \ldots, n$, since $Sv_i \in \operatorname{im} S \subseteq \operatorname{im} T$, there exists $u_i \in V$ such that $Tu_i = Sv_i$. Define a linear map $E \in \operatorname{End}(V)$ by

$$Ev_i = u_i \quad \text{for } i = 1, \ldots, n.$$

Then for any $v \in V$, we can express v as a linear combination of the basis vectors:

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

for some scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Applying S and TE to both sides, we have

$$Sv = \alpha_1 Sv_1 + \cdots + \alpha_n Sv_n = \alpha_1 Tu_1 + \cdots + \alpha_n Tu_n = T(\alpha_1 u_1 + \cdots + \alpha_n u_n) = TEv.$$

Therefore, $S = TE$.

Conversely, suppose there exists $E \in \operatorname{End}(V)$ such that $S = TE$. Let $w \in \operatorname{im} S$. Then there exists $v \in V$ such that $Sv = w$. Applying the definition of S, we have

$$w = Sv = TE(v) = T(E(v)).$$

Therefore, $w \in \operatorname{im} T$, and hence $\operatorname{im} S \subseteq \operatorname{im} T$. □

Solution 3B.27 Let $v \in V$. We can express v as

$$v = (v - Pv) + Pv.$$

Observe that $Pv \in \operatorname{im} P$ by definition. Now, we need to show that $v - Pv \in \ker P$. Applying P to $v - Pv$, we have

$$P(v - Pv) = Pv - P^2v = Pv - Pv = 0.$$

Therefore, $v - Pv \in \ker P$. This shows that every vector $v \in V$ can be expressed as a sum of a vector in $\ker P$ and a vector in $\operatorname{im} P$, which implies that

$$V = \ker P + \operatorname{im} P.$$

Next, we need to show that the sum is direct. Suppose $u \in \ker P \cap \operatorname{im} P$. Then there exists some $w \in V$ such that $u = Pw$. Since $u \in \ker P$, we have

$$Pu = P(Pw) = P^2w = Pw = u.$$

But since $u \in \ker P$, we also have $Pu = 0$. Therefore, we have

$$u = Pu = 0.$$

This shows that the intersection of $\ker P$ and $\operatorname{im} P$ is trivial, i.e., $\ker P \cap \operatorname{im} P = \{0\}$. Hence, the sum is direct, and we conclude that

$$V = \ker P \oplus \operatorname{im} P. \qquad \square$$

Solution 3B.28

Manual Remark. *Note that* $\mathbb{R}[X]$ *is infinite-dimensional, so we cannot use the fundamental theorem of linear maps directly. Instead, one should consider finite-dimensional subspaces of* $\mathbb{R}[X]$.

If $p = 0$, then $p = D(0) \in \operatorname{im} D$ by linearity. Suppose $p \neq 0$ and let $n = \deg p$. Consider the list $D(x), D(x^2), \ldots, D(x^{n+1})$. Since x^k is nonconstant for each $k \in \{1, \ldots, n+1\}$, we have $\deg D(x^k) = k - 1$. Thus, $D(x), D(x^2), \ldots, D(x^{n+1})$ is a list of $n + 1$ polynomials with pairwise distinct degrees, which implies linear independence. Since $\dim \mathbb{R}[X]_{\leq n} = n + 1$, this list is a basis for $\mathbb{R}[X]_{\leq n}$. In particular, $p \in \mathbb{R}[X]_{\leq n}$, so there exist scalars $c_1, \ldots, c_{n+1} \in \mathbb{R}$ such that

$$p = \sum_{k=1}^{n+1} c_k D(x^k) = D\left(\sum_{k=1}^{n+1} c_k x^k\right) \in \operatorname{im} D.$$

Since p was arbitrary, D is surjective. $\square$

Solution 3B.29

Appendix. *An alternate proof is given at the end of this solution.*

Define $D \in \mathrm{End}(\mathbb{R}[X])$ by $Dq = 5q'' + 3q'$ for all $q \in \mathbb{R}[X]$. For nonconstant q, we have $\deg Dq = \deg q - 1$. By Exercise 3B28, D is surjective. Therefore, for any polynomial $p \in \mathbb{R}[X]$, there exists a polynomial $q \in \mathbb{R}[X]$ such that $Dq = p$, which means that $5q'' + 3q' = p$.

Alternate proof. Coolempire93. Let $p(x) = a_0 + a_1x + \cdots + a_nx^n$ and suppose there exists a polynomial $q(x) = b_0 + b_1x + \cdots + b_mx^m$ satisfying $5q'' + 3q' = p$. If $\deg q = m \geq 2$, we can say q' has degree $m-1$ and q'' has degree $m-2$, so $5q'' + 3q'$ has degree $m-1$. To satisfy the given equation, then, we must have $m-1 = n = \deg p$, and we may write $q(x) = b_0 + b_1x + \cdots + b_{n+1}x^{n+1}$. Before we deal with the general case, we will quickly consider when $n = 0$ (which would force $m < 2$):

If $p(x) = a_0$, then let $q(x) = b_0 + b_1x$. Because $5q'' + 3q' = p$, we have $5q'' + 3q' = 5(0) + 3b_1 = a_0$, so we set $b_1 = \frac{a_0}{3}$, and we have a polynomial q such that $5q'' + 3q' = p$ for any value of b_0. As we will see below, the constant term of q never depends on p.

For the remainder of this proof, we take $m = n + 1$ and suppose $n \geq 1$. We have

$$\begin{aligned} q'(x) &= b_1 + 2b_2x + \cdots + (n+1)b_{n+1}x^n \\ q''(x) &= 2b_2 + 6b_3x + \cdots + n(n+1)b_{n+1}x^{n-1} \end{aligned}$$

Equating coefficients of $5q'' + 3q'$ with those of p, we obtain the system

$$\begin{cases} a_0 = 3b_1 + 10b_2 \\ a_1 = 6b_2 + 30b_3 \\ \vdots \\ a_{n-1} = 3nb_n + 5n(n+1)b_{n+1} \\ a_n = 3(n+1)b_{n+1} \end{cases}.$$

That is, $a_i = 3(i+1)b_{i+1} + 5(i+1)(i+2)b_{i+2}$ for $0 \leq i < n$ and $a_n = 3(n+1)b_{n+1}$. Finally, we solve for the coefficients of q by back-substitution. Noting that

$$b_j = \frac{a_{j-1} - 5j(j+1)b_{j+1}}{3j} = \frac{a_{j-1}}{3j} - \frac{5}{3}(j+1)b_{j+1},$$

we have

$$\begin{cases} b_{n+1} = \frac{a_n}{3(n+1)} \\ b_n = \frac{a_{n-1}}{3n} - \frac{5}{3}(n+1)b_{n+1} = \frac{a_{n-1}}{3n} - \frac{5}{9}a_n \\ b_{n-1} = \frac{a_{n-2}}{3(n-1)} - \frac{5}{3}nb_n = \frac{a_{n-2}}{3(n-1)} - \frac{5}{9}a_{n-1} + \frac{25}{27}na_n \\ \vdots \\ b_1 = \frac{a_0 - 10b_2}{3} \end{cases}$$

which has a valid solution[†] since no division by 0 occurs. Finally, we note that b_0 is a free variable, so there is an infinite family of polynomials $q \in \mathbb{R}[X]$ with the given coefficients such that $5q'' + 3q' = p$.

† Solving this recurrence is outside the scope of this course, but, specifically, the coefficients of q are

$$b_j = \frac{a_{j-1}}{3j} - \frac{5}{9}a_j + \frac{25}{27}(j+1)a_{j+1} - \frac{125}{81}(j+1)(j+2)a_{j+2} + \cdots + (-1)^{n-j+1}\frac{5^{n-j+1}}{3^{n-j+2}}(j+1)(j+2)\cdots n a_n$$
$$= \frac{1}{3}\left(\frac{a_{j-1}}{j} + \sum_{i=j}^{n}\left(-\frac{5}{3}\right)^{i-j+1}\frac{i!}{j!}a_i\right).$$

□

Solution 3B.30 Let $v \in V$. We want to show that v can be expressed as a sum of a vector in $\ker\phi$ and a scalar multiple of u. Since $\phi(u) \neq 0$, we can define a scalar

$$a = \frac{\phi(v)}{\phi(u)}.$$

Then consider the vector

$$w = v - au.$$

We have

$$\phi(w) = \phi(v - au) = \phi(v) - a\phi(u) = \phi(v) - \frac{\phi(v)}{\phi(u)}\phi(u) = 0.$$

Therefore, $w \in \ker\phi$. Thus, we can express v as

$$v = w + au,$$

where $w \in \ker\phi$ and au is a scalar multiple of u. This shows that every vector in V can be expressed as a sum of a vector in $\ker\phi$ and a scalar multiple of u, which implies that

$$V = \ker\phi + \{au : a \in \mathbb{F}\}.$$

Next, we need to show that the sum is direct. Suppose $w \in \ker\phi \cap \{au : a \in \mathbb{F}\}$. Then there exists a scalar $a \in \mathbb{F}$ such that $w = au$. Since $w \in \ker\phi$, we have

$$0 = \phi(w) = \phi(au) = a\phi(u).$$

Since $\phi(u) \neq 0$, it follows that $a = 0$. Therefore, $w = 0$, which shows that the intersection of $\ker\phi$ and $\{au : a \in \mathbb{F}\}$ is trivial. Hence, the sum is direct, and we conclude that

$$V = \ker\phi \oplus \{au : a \in \mathbb{F}\}.$$

□

Solution 3B.31 Suppose there exists $T \in \mathrm{Hom}(V, W)$ such that $\ker T = X$ and $\operatorname{im} T = Y$. By the fundamental theorem of linear maps, we have

$$\dim V = \dim\ker T + \dim\operatorname{im} T = \dim X + \dim Y.$$

Conversely, suppose $\dim X + \dim Y = \dim V$. Let $n = \dim V$, $k = \dim X$, and $m = \dim Y$. Then $n = k + m$. Choose a basis $\{x_1, \ldots, x_k\}$ of X and extend it to a

basis $\{x_1, \ldots, x_k, v_1, \ldots, v_m\}$ of V. Let $\{y_1, \ldots, y_m\}$ be a basis of Y. Define a linear map $T \in \mathrm{Hom}(V, W)$ by

$$Tx_i = 0 \quad \text{for } i = 1, \ldots, k,$$

and

$$Tv_j = y_j \quad \text{for } j = 1, \ldots, m.$$

Then for any $v \in V$, we can express v as a linear combination of the basis vectors:

$$v = \alpha_1 x_1 + \cdots + \alpha_k x_k + \beta_1 v_1 + \cdots + \beta_m v_m$$

for some scalars $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_m \in \mathbb{F}$. Applying T to both sides gives

$$Tv = \alpha_1 Tx_1 + \cdots + \alpha_k Tx_k + \beta_1 Tv_1 + \cdots + \beta_m Tv_m = 0 + \beta_1 y_1 + \cdots + \beta_m y_m.$$

Therefore, $Tv \in Y$. Now suppose $y = \gamma_1 y_1 + \cdots + \gamma_m y_m \in Y$. Then

$$T(\gamma_1 v_1 + \cdots + \gamma_m v_m) = \gamma_1 y_1 + \cdots + \gamma_m y_m = y.$$

Hence, $\operatorname{im} T = Y$. Moreover, if $Tv = 0$, then $\beta_1 y_1 + \cdots + \beta_m y_m = 0$. Since $y_1, \ldots, y_m$ are linearly independent, it follows that $\beta_1 = \cdots = \beta_m = 0$. Thus, $v \in X$. Additionally, if $v \in X$, then $v = \alpha_1 x_1 + \cdots + \alpha_k x_k$ for some scalars $\alpha_1, \ldots, \alpha_k \in \mathbb{F}$, and hence $Tv = 0$. Therefore, $\ker T = X$. This shows that there exists a linear map $T \in \mathrm{Hom}(V, W)$ such that $\ker T = X$ and $\operatorname{im} T = Y$. □

Solution 3B.32 Observe that $\ker \phi$ is a two-sided ideal of $\mathrm{End}(V)$ as follows: if $X \in \ker \phi$ and $Y \in \mathrm{End}(V)$, then

$$\phi(XY) = \phi(X)\phi(Y) = 0 \cdot \phi(Y) = 0, \quad \phi(YX) = \phi(Y)\phi(X) = \phi(Y) \cdot 0 = 0.$$

Since $\dim V > 1$, then Exercise 3A16 gives the existence of $S, T \in \mathrm{End}(V)$ such that $ST \neq TS$. Then we have

$$\phi(ST - TS) = \phi(S)\phi(T) - \phi(T)\phi(S) = 0.$$

Therefore, $ST - TS \in \ker \phi$. Since $\ker \phi$ is a two-sided ideal of $\mathrm{End}(V)$, then by Exercise 3A17, we have $\ker \phi = \mathrm{End}(V)$. Hence, $\phi = 0$. □

Solution 3B.33

(a) Let $u + vi$ and $w + xi \in V_{\mathbb{C}}$. Then for any $a + bi \in \mathbb{C}$, we have

$$\begin{aligned} T_{\mathbb{C}}((a+bi)(u+vi) + (w+xi)) &= T_{\mathbb{C}}((au - bv + (av + bu)i) + (w + xi)) \\ &= T(au - bv + w) + iT(av + bu + x) \\ &= aTu - bTv + Tw + i(aTv + bTu + Tx) \\ &= (a+bi)(Tu + iTv) + (Tw + iTx) \\ &= (a+bi)T_{\mathbb{C}}(u+vi) + T_{\mathbb{C}}(w+xi). \end{aligned}$$

Therefore, $T_{\mathbb{C}}$ is a complex linear map from $V_{\mathbb{C}}$ to $W_{\mathbb{C}}$.

(b) Suppose $T_{\mathbb{C}}$ is injective. Let $v \in V$ be such that $Tv = 0$. Then

$$T_{\mathbb{C}}(v + i0) = Tv + iT0 = 0 + i0 = 0.$$

Since $T_{\mathbb{C}}$ is injective, we have $v + i0 = 0$, which implies that $v = 0$. Therefore, T is injective.

Conversely, suppose T is injective. Let $u + iv \in V_{\mathbb{C}}$ be such that $T_{\mathbb{C}}(u + iv) = 0$. Then

$$Tu + iTv = 0.$$

This implies that $Tu = 0$ and $Tv = 0$. Since T is injective, we have $u = 0$ and $v = 0$. Therefore, $u + iv = 0$, and hence $T_{\mathbb{C}}$ is injective.

(c) Suppose $\operatorname{im} T_{\mathbb{C}} = W_{\mathbb{C}}$. Let $w \in W$. Then $w + i0 \in W_{\mathbb{C}}$, and since $\operatorname{im} T_{\mathbb{C}} = W_{\mathbb{C}}$, there exists $u + iv \in V_{\mathbb{C}}$ such that

$$T_{\mathbb{C}}(u + iv) = Tu + iTv = w + i0.$$

This implies that $Tu = w$ and $Tv = 0$. Therefore, $w \in \operatorname{im} T$, and hence $\operatorname{im} T = W$.

Conversely, suppose $\operatorname{im} T = W$. Let $w_1 + iw_2 \in W_{\mathbb{C}}$. Since $\operatorname{im} T = W$, there exist $u, v \in V$ such that $Tu = w_1$ and $Tv = w_2$. Then

$$T_{\mathbb{C}}(u + iv) = Tu + iTv = w_1 + iw_2.$$

Therefore, $w_1 + iw_2 \in \operatorname{im} T_{\mathbb{C}}$, and hence $\operatorname{im} T_{\mathbb{C}} = W_{\mathbb{C}}$.

□

§3C

Solution 3C.1 Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_m$ be a basis of W. Let $A = \operatorname{Mat}(T)$ be the matrix with respect to these bases. Then the j-th column of A is the coordinate vector of Tv_j with respect to the basis for W.
Suppose A had fewer than $\dim \operatorname{im} T$ nonzero columns. Since zero columns do not contribute to the span, the columns of A would then span a space of dimension less than $\dim \operatorname{im} T$. This is a contradiction, since fewer than $\dim \operatorname{im} T$ vectors cannot span $\operatorname{im} T$. Therefore, A must have at least $\dim \operatorname{im} T$ nonzero columns, and hence at least $\dim \operatorname{im} T$ nonzero entries. □

Solution 3C.2 Suppose $\dim \operatorname{im} T = 1$. Then there exists a nonzero vector $y_1 \in W$ such that $\operatorname{im} T = \operatorname{span} y_1$. Let $\dim V = n$ and $\dim W = m$. Extend y_1 to a basis $y_1, \ldots, y_m$ of W, and define the following vectors in W:

$$w_1 = y_1 - (y_2 + \cdots + y_m), \quad w_k = y_k \text{ for } k = 2, \ldots, m.$$

It is clear that $w_1 + w_2 + \cdots + w_m = y_1$, and hence $w_1, \ldots, w_m$ spans W. Now suppose for scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{F}$, we have

$$\alpha_1 w_1 + \alpha_2 w_2 + \cdots + \alpha_m w_m = 0.$$

Then we can express this as

$$\alpha_1(y_1 - (y_2 + \cdots + y_m)) + \alpha_2 y_2 + \cdots + \alpha_m y_m = 0.$$

Simplifying, we get

$$(\alpha_1)y_1 + (-\alpha_1 + \alpha_2)y_2 + \cdots + (-\alpha_1 + \alpha_m)y_m = 0.$$

Since $y_1, \ldots, y_m$ are linearly independent, we must have $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$. Therefore, $w_1, \ldots, w_m$ are linearly independent and hence form a basis of W.
We now construct a basis of V. By the fundamental theorem of linear maps, we have

$$\dim V = \dim T + \dim \operatorname{im} T = \dim T + 1.$$

Let $u_2, \ldots, u_n$ be a basis of T. Then we can extend this to a basis $u_1, u_2, \ldots, u_n$ of V, where $u_1 \notin T$. Since $\operatorname{im} T = \operatorname{span} y_1$, we have

$$Tu_1 = \beta y_1$$

for some nonzero scalar $\beta \in \mathbb{F}$. If $\beta \neq 1$, we may replace u_1 with $\beta^{-1}u_1$ to ensure that $Tu_1 = y_1$. We construct the following vectors in V:

$$v_1 = u_1, \quad v_k = u_1 + u_k \text{ for } k = 2, \ldots, n.$$

Note that for every $j = 1, \ldots, n$, we have $Tv_j = Tu_1 + Tu_j = y_1 + 0 = y_1$. We claim that $v_1, \ldots, v_n$ form a basis of V. Suppose for scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, we have

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0.$$

Then we can express this as

$$(\alpha_1 + \alpha_2 + \cdots + \alpha_n)u_1 + \alpha_2 u_2 + \cdots + \alpha_n u_n = 0.$$

Since $u_1, \ldots, u_n$ are linearly independent, we must have $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 0$ and $\alpha_2 = \cdots = \alpha_n = 0$. This implies that $\alpha_1 = 0$, and hence $v_1, \ldots, v_n$ are linearly independent and form a basis of V. Then with respect to the bases $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_m$ of W, we have that $Tv_j = y_1 = w_1 + w_2 + \cdots + w_m$ for each $j = 1, \ldots, n$. Therefore, the matrix of T with respect to these bases has all entries equal to 1.

Conversely, suppose there exist a basis of V and a basis of W such that with respect to these bases, all entries of $\operatorname{Mat}(T)$ equal 1. Then for each basis vector v_j of V, we have

$$Tv_j = w$$

for some fixed nonzero vector $w \in W$, where $w = w_1 + \cdots + w_m$ is the sum of the basis vectors of W. Since $\operatorname{im} T$ is spanned by the images of the basis vectors of V, then any sum of Tv_j is a scalar multiple of w. Therefore, $\operatorname{im} T = \operatorname{span} w$, and hence $\dim \operatorname{im} T = 1$. □

Solution 3C.3

(a) Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_m$ be a basis of W. Then the matrix of $S + T$ with respect to these bases is given by

$$\begin{aligned}\operatorname{Mat}(S+T) &= \begin{pmatrix}(S+T)v_1 & (S+T)v_2 & \cdots & (S+T)v_n\end{pmatrix} \\ &= \begin{pmatrix}Sv_1 + Tv_1 & Sv_2 + Tv_2 & \cdots & Sv_n + Tv_n\end{pmatrix} \\ &= \begin{pmatrix}Sv_1 & Sv_2 & \cdots & Sv_n\end{pmatrix} + \begin{pmatrix}Tv_1 & Tv_2 & \cdots & Tv_n\end{pmatrix} \\ &= \operatorname{Mat}(S) + \operatorname{Mat}(T).\end{aligned}$$

(b) Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_m$ be a basis of W. Then the matrix of λT with respect to these bases is given by

$$\begin{aligned}\operatorname{Mat}(\lambda T) &= \begin{pmatrix}(\lambda T)v_1 & (\lambda T)v_2 & \cdots & (\lambda T)v_n\end{pmatrix} \\ &= \begin{pmatrix}\lambda Tv_1 & \lambda Tv_2 & \cdots & \lambda Tv_n\end{pmatrix} \\ &= \lambda \begin{pmatrix}Tv_1 & Tv_2 & \cdots & Tv_n\end{pmatrix} \\ &= \lambda \operatorname{Mat}(T).\end{aligned}$$

□

Solution 3C.4 Consider the standard basis of $\mathbb{R}[X]_{\leq 3}$. Reorder it to obtain the basis $x^3, x^2, x, 1$. Consider the standard basis of $\mathbb{R}[X]_{\leq 2}$. Reorder it to obtain the basis $x^2, x, 1$. Observe that $D(x^3) = 3x^2$, $D(x^2) = 2x$, $D(x) = 1$, and $D(1) = 0$. However, we need the matrix of D to have the desired form. To achieve this, we can scale the basis vectors of $\mathbb{R}[X]_{\leq 3}$ and $\mathbb{R}[X]_{\leq 2}$ appropriately. Then we may define the basis for $\mathbb{R}[X]_{\leq 3}$ as

$$\frac{1}{3}x^3, \frac{1}{2}x^2, x, 1$$

while maintaining the standard basis for $\mathbb{R}[X]_{\leq 2}$ as $x^2, x, 1$. Then we have

$$D\left(\frac{1}{3}x^3\right) = x^2, \quad D\left(\frac{1}{2}x^2\right) = x, \quad D(x) = 1, \quad D(1) = 0.$$

Therefore, with respect to the bases, we get the desired matrix representation of D. □

Solution 3C.5 Let $n = \dim V$ and $m = \dim W$. By the fundamental theorem of linear maps, we have

$$\dim V = \dim \ker T + \dim \operatorname{im} T.$$

Let $r = \dim \operatorname{im} T$. Let $u_1, \ldots, u_{n-r}$ be a basis of $\ker T$. Extend this to a basis $v_1, \ldots, v_r, u_1, \ldots, u_{n-r}$ of V. We claim that $Tv_1, \ldots, Tv_r$ are linearly independent. Suppose for scalars $\alpha_1, \ldots, \alpha_r \in \mathbb{F}$, we have

$$\alpha_1 Tv_1 + \cdots + \alpha_r Tv_r = 0.$$

Then

$$T(\alpha_1 v_1 + \cdots + \alpha_r v_r) = 0,$$

which implies that $\alpha_1 v_1 + \cdots + \alpha_r v_r \in \ker T$. Since $v_1, \ldots, v_r, u_1, \ldots, u_{n-r}$ is a basis of V, it follows that $\alpha_1 = \cdots = \alpha_r = 0$. Therefore, $Tv_1, \ldots, Tv_r$ are linearly independent

and form a basis of $\operatorname{im} T$. Extend this to a basis $Tv_1, \ldots, Tv_r, w_1, \ldots, w_{m-r}$ of W. Then with respect to these bases, Tv_k is the k-th basis vector of W for each $k = 1, \ldots, r$, and

$$Tu_j = 0, \quad \text{for } j = 1, \ldots, n-r.$$

Therefore, the matrix of T with respect to these bases has 1's in the diagonal entries corresponding to Tv_k and 0's elsewhere, as desired. □

Solution 3C.6 Let $n = \dim W$. If $Tv_1 = 0$, then we can choose any basis of W and the first column of $\operatorname{Mat}(T)$ will be all zeros. If $Tv_1 \neq 0$, then we can extend Tv_1 to a basis of W. Let $w_1 = Tv_1$ and extend it to a basis $w_1, w_2, \ldots, w_n$ of W. Then with respect to the bases $v_1, \ldots, v_m$ of V and $w_1, \ldots, w_n$ of W, we have

$$Tv_1 = w_1$$

and for each $j = 2, \ldots, m$, we can express

$$Tv_j = \alpha_{j1}w_1 + \alpha_{j2}w_2 + \cdots + \alpha_{jn}w_n$$

for some scalars $\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jn} \in \mathbb{F}$. Therefore, the first column of $\operatorname{Mat}(T)$ has a 1 in the first row and zeros in all other rows, as desired. □

Solution 3C.7 Let $m = \dim V$ and $u_1, \ldots, u_m$ be a basis for V. If T has trivial w_1 component, meaning that the coefficient of w_1 in Tv is zero for all $v \in V$, then choose any basis of V and the first row of $\operatorname{Mat}(T)$ will be all zeros. Otherwise, reorder the basis so that Tu_1 has non-trivial w_1 component. Then we can express

$$Tu_1 = \alpha w_1 + \beta_2 w_2 + \cdots + \beta_n w_n$$

for some scalars $\alpha, \beta_2, \ldots, \beta_n \in \mathbb{F}$ such that $\alpha \neq 0$. We can scale u_1 by α^{-1} to obtain a new vector $v_1 = \alpha^{-1}u_1$ such that

$$Tv_1 = w_1 + \frac{\beta_2}{\alpha}w_2 + \cdots + \frac{\beta_n}{\alpha}w_n.$$

With the remaining $u_2, \ldots, u_m$ basis vectors of V, note that they can be expressed as

$$Tu_j = \alpha_{j1}w_1 + \alpha_{j2}w_2 + \cdots + \alpha_{jn}w_n$$

for some scalars $\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jn} \in \mathbb{F}$. Consider the set of vectors of V given by

$$v_1 = \alpha^{-1}u_1, \quad v_j = u_j - \alpha_{j1}v_1 \text{ for } j = 2, \ldots, m.$$

Note that $Tv_j = Tu_j - \alpha_{j1}Tv_1$ is just a combination of $w_2, \ldots, w_n$. We now claim that $v_1, v_2, \ldots, v_m$ are linearly independent. Suppose for scalars $\gamma_1, \ldots, \gamma_m \in \mathbb{F}$, we have

$$\gamma_1 v_1 + \gamma_2 v_2 + \cdots + \gamma_m v_m = 0.$$

Then we can express this as

$$\gamma_1 v_1 + \gamma_2(u_2 - \alpha_{21}v_1) + \cdots + \gamma_m(u_m - \alpha_{m1}v_1) = 0.$$

Simplifying, we get

$$(\gamma_1 - \gamma_2\alpha_{21} - \cdots - \gamma_m\alpha_{m1})v_1 + \gamma_2 u_2 + \cdots + \gamma_m u_m = 0.$$

Since $v_1, u_2, \ldots, u_m$ are linearly independent, we must have $\gamma_1 - \gamma_2\alpha_{21} - \cdots - \gamma_m\alpha_{m1} = 0$ and $\gamma_2 = \cdots = \gamma_m = 0$. This implies that $\gamma_1 = 0$, and hence $v_1, v_2, \ldots, v_m$ are linearly independent. Then $v_1, v_2, \ldots, v_m$ form a basis of V. It follows that Tv_1 has a 1 in the first row, while Tv_j for $j = 2, \ldots, m$ have zeros in the first row. □

Solution 3C.8 Let A be an m-by-n matrix and B be an n-by-p matrix. The entry in row j, column k, of the product AB is given by

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r} B_{r,k}.$$

The row vector $A_{j,\cdot}$ is the j-th row of A, which can be expressed as

$$A_{j,\cdot} = (A_{j,1}, A_{j,2}, \ldots, A_{j,n}).$$

When we multiply this row vector by the matrix B, we get a new row vector whose entries are given by

$$(A_{j,\cdot}B)_k = \sum_{r=1}^{n} A_{j,r} B_{r,k}.$$

Comparing this with the expression for $(AB)_{j,k}$, we see that

$$(AB)_{j,k} = (A_{j,\cdot}B)_k$$

for each column index k. Therefore, the entire row vector $(AB)_{j,\cdot}$ is equal to the row vector $A_{j,\cdot}B$. This proves that row j of AB equals (row j of A) times B. □

Solution 3C.9 Let $a = (a_1 \quad \cdots \quad a_n)$ be a 1-by-n matrix and B be an n-by-p matrix. The product aB is a 1-by-p matrix, and its entries are given by

$$(aB)_{1,k} = \sum_{j=1}^{n} a_j B_{j,k}$$

for each column index k.
Now, consider the linear combination of the rows of B given by

$$a_1 B_{1,\cdot} + a_2 B_{2,\cdot} + \cdots + a_n B_{n,\cdot}.$$

The entry in row 1, column k, of this linear combination is

$$(a_1 B_{1,\cdot} + a_2 B_{2,\cdot} + \cdots + a_n B_{n,\cdot})_k = a_1 B_{1,k} + a_2 B_{2,k} + \cdots + a_n B_{n,k} = \sum_{j=1}^{n} a_j B_{j,k}.$$

Comparing this with the expression for $(aB)_{1,k}$, we see that

$$(aB)_{1,k} = (a_1 B_{1,\cdot} + a_2 B_{2,\cdot} + \cdots + a_n B_{n,\cdot})_k$$

for each column index k. Therefore, we conclude that

$$aB = a_1 B_{1,\cdot} + a_2 B_{2,\cdot} + \cdots + a_n B_{n,\cdot},$$

which shows that aB is indeed a linear combination of the rows of B, with the scalars coming from the entries of a. □

Solution 3C.10 Let

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then we compute the products AB and BA:

$$AB = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 \cdot 0 + 2 \cdot 1 & 1 \cdot 1 + 2 \cdot 0 \\ 3 \cdot 0 + 4 \cdot 1 & 3 \cdot 1 + 4 \cdot 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix},$$

and

$$BA = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 0 \cdot 1 + 1 \cdot 3 & 0 \cdot 2 + 1 \cdot 4 \\ 1 \cdot 1 + 0 \cdot 3 & 1 \cdot 2 + 0 \cdot 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

By comparing the two products, we see that $AB \neq BA$. □

Solution 3C.11 Let A be an m-by-n matrix, and let B and C be n-by-p matrices. Then $B+C$ is an n-by-p matrix, and hence $A(B+C)$ is an m-by-p matrix. The products AB and AC are also defined, and both are m-by-p matrices. Therefore, the sum $AB + AC$ is defined and is also an m-by-p matrix. By Exercise 3C8 and Exercise 3C9, we have

$$A(B + C)_{j,\cdot} = A_{j,\cdot}(B + C) = A_{j,\cdot}B + A_{j,\cdot}C = (AB)_{j,\cdot} + (AC)_{j,\cdot} = (AB + AC)_{j,\cdot}$$

for all $j = 1, \ldots, m$. Therefore, $A(B + C) = AB + AC$.
Similarly, let D and E be m-by-n matrices, and let F be an n-by-p matrix. Then $D+E$ is an m-by-n matrix, and hence $(D + E)F$ is an m-by-p matrix. The products DF and EF are also defined, and both are m-by-p matrices. Therefore, the sum $DF + EF$ is defined and is also an m-by-p matrix. Then the product $(D + E)F$ is defined. By Exercise 3C8 and Exercise 3C9, we have

$$(D + E)F_{j,\cdot} = (D + E)_{j,\cdot}F = D_{j,\cdot}F + E_{j,\cdot}F = (DF)_{j,\cdot} + (EF)_{j,\cdot} = (DF + EF)_{j,\cdot}$$

for all $j = 1, \ldots, m$. Therefore, $(D + E)F = DF + EF$. □

Solution 3C.12

Manual Remark. *Unfortunately, we cannot use the same trick as in Exercise 3C11 to prove this result.*

Let A be an m-by-n matrix, B be an n-by-p matrix, and C be a p-by-q matrix. Then the product AB is defined and is an m-by-p matrix. Therefore, the product $(AB)C$ is also defined and is an m-by-q matrix. Going by the definition of matrix multiplication, we have

$$((AB)C)_{j,k} = \sum_{r=1}^{p} (AB)_{j,r} C_{r,k} = \sum_{r=1}^{p} \left(\sum_{s=1}^{n} A_{j,s} B_{s,r} \right) C_{r,k} = \sum_{r=1}^{p} \sum_{s=1}^{n} A_{j,s} B_{s,r} C_{r,k}.$$

On the other hand, the product BC is defined and is an n-by-q matrix. Therefore, the product $A(BC)$ is also defined and is an m-by-q matrix. Again, by the definition of matrix multiplication, we have

$$(A(BC))_{j,k} = \sum_{s=1}^{n} A_{j,s}(BC)_{s,k} = \sum_{s=1}^{n} A_{j,s}\left(\sum_{r=1}^{p} B_{s,r}C_{r,k}\right) = \sum_{s=1}^{n}\sum_{r=1}^{p} A_{j,s}B_{s,r}C_{r,k}.$$

Since the order of summation does not matter, we can interchange the order of the sums in the last expression to get

$$(A(BC))_{j,k} = \sum_{r=1}^{p}\sum_{s=1}^{n} A_{j,s}B_{s,r}C_{r,k}.$$

Comparing the expressions for $((AB)C)_{j,k}$ and $(A(BC))_{j,k}$, we see that they are equal for all j and k. Therefore, $(AB)C = A(BC)$. □

Solution 3C.13 Note that $A^3 = (AA)A = A^2A$. By the definition of matrix multiplication, we have

$$(A^3)_{j,k} = (A^2A)_{j,k} = \sum_{p=1}^{n}(A^2)_{j,p}A_{p,k}.$$

Now, we can express $(A^2)_{j,p}$ as

$$(A^2)_{j,p} = (AA)_{j,p} = \sum_{r=1}^{n} A_{j,r}A_{r,p}.$$

Substituting this back into the expression for $(A^3)_{j,k}$, we get

$$(A^3)_{j,k} = \sum_{p=1}^{n}\left(\sum_{r=1}^{n} A_{j,r}A_{r,p}\right)A_{p,k} = \sum_{p=1}^{n}\sum_{r=1}^{n} A_{j,r}A_{r,p}A_{p,k}.$$

By renaming the indices r and p, we can rewrite this as

$$(A^3)_{j,k} = \sum_{p=1}^{n}\sum_{r=1}^{n} A_{j,p}A_{p,r}A_{r,k},$$

which is the desired expression. □

Solution 3C.14 Let $A, B \in \mathbb{F}^{m,n}$ and let $\lambda \in \mathbb{F}$. We need to show that $(A+B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}}$ and $(\lambda A)^{\mathrm{t}} = \lambda A^{\mathrm{t}}$.
First, consider the sum $A + B$. The entry in row j, column k, of $A + B$ is given by

$$(A+B)_{j,k} = A_{j,k} + B_{j,k}.$$

Taking the transpose, we have

$$((A+B)^{\mathrm{t}})_{k,j} = (A+B)_{j,k} = A_{j,k} + B_{j,k} = (A^{\mathrm{t}})_{k,j} + (B^{\mathrm{t}})_{k,j} = (A^{\mathrm{t}} + B^{\mathrm{t}})_{k,j}.$$

Therefore, $(A+B)^{\mathrm{t}} = A^{\mathrm{t}} + B^{\mathrm{t}}$.

Next, consider the scalar multiplication λA. The entry in row j, column k, of λA is given by

$$(\lambda A)_{j,k} = \lambda A_{j,k}.$$

Taking the transpose, we have

$$((\lambda A)^{\mathrm{t}})_{k,j} = (\lambda A)_{j,k} = \lambda A_{j,k} = \lambda (A^{\mathrm{t}})_{k,j} = (\lambda A^{\mathrm{t}})_{k,j}.$$

Therefore, $(\lambda A)^{\mathrm{t}} = \lambda A^{\mathrm{t}}$.

Since both properties hold, we conclude that the function $A \mapsto A^{\mathrm{t}}$ is a linear map from $\mathbb{F}^{m,n}$ to $\mathbb{F}^{n,m}$. □

Solution 3C.15 Let A be an m-by-n matrix and C be an n-by-p matrix. The entry in row j, column k, of the product AC is given by

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Taking the transpose, we have

$$((AC)^{\mathrm{t}})_{k,j} = (AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Now, consider the product $C^{\mathrm{t}} A^{\mathrm{t}}$. The entry in row k, column j, of this product is given by

$$(C^{\mathrm{t}} A^{\mathrm{t}})_{k,j} = \sum_{r=1}^{n} (C^{\mathrm{t}})_{k,r} (A^{\mathrm{t}})_{r,j} = \sum_{r=1}^{n} C_{r,k} A_{j,r}.$$

Since multiplication in the field $\mathbb{F}$ is commutative, we can rewrite this as

$$(C^{\mathrm{t}} A^{\mathrm{t}})_{k,j} = \sum_{r=1}^{n} A_{j,r} C_{r,k},$$

which is exactly the same as the expression for $((AC)^{\mathrm{t}})_{k,j}$.

Therefore, we conclude that

$$(AC)^{\mathrm{t}} = C^{\mathrm{t}} A^{\mathrm{t}}.$$

□

Solution 3C.16 Suppose the rank of A is 1. Then the column rank of A is 1, which means that all columns of A are scalar multiples of a single non-zero column vector. Let $c = (c_1, \ldots, c_m)^{\mathrm{t}}$ be this non-zero column vector. Then for each column k of A, there exists a scalar $d_k \in \mathbb{F}$ such that the k-th column of A is given by $d_k c$. Therefore, we can express the entries of A as

$$A_{j,k} = c_j d_k$$

for all $j = 1, \ldots, m$ and $k = 1, \ldots, n$.

Conversely, suppose there exist $(c_1, \ldots, c_m) \in \mathbb{F}^m$ and $(d_1, \ldots, d_n) \in \mathbb{F}^n$ such that $A_{j,k} = c_j d_k$ for all j and k. Since $A \neq 0$, then $(c_1, \ldots, c_m) \neq 0$ and $(d_1, \ldots, d_n) \neq 0$. This means that all columns of A are scalar multiples of the column vector c, and hence the column rank of A is 1. Therefore, the rank of A is 1. □

Solution 3C.17

- (a) $\iff$ (b): (a) $\iff$ $\dim\ker T = 0$ $\iff$ $\dim\operatorname{im} T = n$ $\iff$ column rank $\operatorname{Mat}(T) = n$. For n column vectors, rank n implies linear independence $\iff$ (b).
- (b) $\iff$ (c): n columns of $\operatorname{Mat}(T)$ are linearly independent in $\mathbb{F}^{n,1}$ $\iff$ n columns of $\operatorname{Mat}(T)$ span $\mathbb{F}^{n,1}$.
- (c) $\iff$ (d): columns span $\mathbb{F}^{n,1}$ $\iff$ rank $\operatorname{Mat}(T) = n$ by (b) $\iff$ (c). By Theorem 3.57, rank $\operatorname{Mat}(T)$ = row rank $\operatorname{Mat}(T)$ $\iff$ rows span $\mathbb{F}^{1,n}$.
- (d) $\iff$ (e): n rows of $\operatorname{Mat}(T)$ span $\mathbb{F}^{1,n}$ $\iff$ n rows of $\operatorname{Mat}(T)$ are linearly independent in $\mathbb{F}^{1,n}$.

□

§3D

Solution 3D.1 Since T is invertible, then $T^{-1} \in \operatorname{Hom}(W, V)$. By $TT^{-1} = I$ and $T^{-1}T = I$, we have that T^{-1} is invertible and $(T^{-1})^{-1} = T$. □

Solution 3D.2 Since T and S are invertible, then $T^{-1} \in \operatorname{Hom}(V, U)$ and $S^{-1} \in \operatorname{Hom}(W, V)$. By the definition of the inverse, we have that

$$(ST)(T^{-1}S^{-1}) = S(TT^{-1})S^{-1} = SIS^{-1} = SS^{-1} = I,$$

and

$$(T^{-1}S^{-1})(ST) = T^{-1}(S^{-1}S)T = T^{-1}IT = T^{-1}T = I.$$

Thus, ST is invertible and $(ST)^{-1} = T^{-1}S^{-1}$. □

Solution 3D.3

- (a) $\implies$ (b): Suppose T is invertible. Consider scalars $\alpha_1, \ldots, \alpha_n$ such that

  $$\alpha_1 Tv_1 + \cdots + \alpha_n Tv_n = 0.$$

  Applying T^{-1} to both sides, we get

  $$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0.$$

  Since $v_1, \ldots, v_n$ is a basis of V, it follows that $\alpha_1 = \cdots = \alpha_n = 0$. Hence, $Tv_1, \ldots, Tv_n$ is linearly independent. Since $Tv_1, \ldots, Tv_n$ has n elements and V is n-dimensional, it follows that $Tv_1, \ldots, Tv_n$ is a basis of V.
- (b) $\implies$ (c): This is trivial.
- (c) $\implies$ (a): Suppose $Tv_1, \ldots, Tv_n$ is a basis of V for some basis $v_1, \ldots, v_n$ of V. Then $\operatorname{im} T = V$. By the fundamental theorem of linear maps, we have that $\dim\ker T = 0$. Thus, T is injective. Since T is injective and $\operatorname{im} T = V$, then T is invertible.

□

Solution 3D.4 Let $v_1, \ldots, v_n$ be a basis of V. Define $T, S \in \mathrm{End}(V)$ such that

$$Tv_1 = 0, \quad Tv_k = v_k \text{ for } k = 2, \ldots, n,$$

and

$$Sv_1 = v_1, \quad Sv_k = 0 \text{ for } k = 2, \ldots, n.$$

Then T and S are non-invertible linear maps, since $\ker T = \operatorname{span} v_1$ and $\ker S = \operatorname{span}(v_2, \ldots, v_n)$. However, $T + S$ is invertible, since

$$(T + S)v_1 = v_1, \quad (T + S)v_k = v_k \text{ for } k = 2, \ldots, n.$$

Hence, $T + S = I$ is invertible. Therefore, the set of non-invertible linear maps from V to itself is not a subspace of $\mathrm{End}(V)$. □

Solution 3D.5 Suppose there exists an invertible linear map T from V to itself such that $Tu = Su$ for every $u \in U$. If $Su = 0$, then $Tu = 0$, and since T is invertible, $u = 0$. Hence, S is injective.
Conversely, suppose S is injective. Let $u_1, \ldots, u_m$ be a basis of U. Since S is injective, then $Su_1, \ldots, Su_m$ is linearly independent. Extend $u_1, \ldots, u_m$ to a basis $u_1, \ldots, u_n$ of V. Extend $Su_1, \ldots, Su_m$ to a basis $v_1, \ldots, v_n$ of V. Define $T \in \mathrm{End}(V)$ such that

$$Tu_k = Su_k \text{ for } k = 1, \ldots, m, \quad Tu_k = v_k \text{ for } k = m + 1, \ldots, n.$$

Then T is invertible, since $Tu_1, \ldots, Tu_n$ is a basis of V. Thus, there exists an invertible linear map T from V to itself such that $Tu = Su$ for every $u \in U$. □

Solution 3D.6 Suppose $\ker S = \ker T$. Define a linear map E_0 from $\operatorname{im} T$ to $\operatorname{im} S$ by setting $E_0(Tv) = Sv$ for all $v \in V$. This is well-defined because if $Tv = Tv'$ for some $v, v' \in V$, then $T(v - v') = 0$, so $v - v' \in \ker T = \ker S$, and thus $S(v - v') = 0$, which implies $Sv = Sv'$. Moreover, if $E_0(Tv) = 0$ for some $v \in V$, then $Sv = 0$, which implies $v \in \ker S = \ker T$, and thus $Tv = 0$. Hence, E_0 is injective. By Exercise 3D5, we can extend E_0 to an invertible linear map E on all of W, and then we have $S = ET$.
Conversely, suppose there exists an invertible $E \in \mathrm{End}(W)$ such that $S = ET$. If $v \in \ker S$, then $Sv = 0$, and since $S = ET$, we have $ETv = 0$. Since E is invertible, $Tv = 0$ so that $v \in \ker T$. Hence, $\ker S \subseteq \ker T$. Similarly, if $v \in \ker T$, then $Tv = 0$, and since $S = ET$, we have $Sv = ETv = E0 = 0$. Thus, $v \in \ker S$, and $\ker T \subseteq \ker S$. Therefore, $\ker S = \ker T$. □

Solution 3D.7 Suppose $\operatorname{im} S = \operatorname{im} T$. Then $\dim \ker S = \dim \ker T$. Let $u_1, \ldots, u_k$ be a basis of $\ker S$ and $v_1, \ldots, v_k$ be a basis of $\ker T$. Extend both bases to a basis of V, say $u_1, \ldots, u_n$ and $v_1, \ldots, v_n$. Observe that for $i > k$, we have that $Su_i \in \operatorname{im} S = \operatorname{im} T$. Then there must exist $y_i \in V$ such that $Ty_i = Su_i$. Define a linear map $E \in \mathrm{End}(V)$ by setting $Eu_i = v_i$ for $i \leq k$ and $Eu_i = y_i$ for $i > k$. To see that $TE = S$, we have for $i \leq k$ that $TEu_i = Tv_i = 0 = Su_i$. Moreover, for $i > k$, we have $TEu_i = Ty_i = Su_i$. To see that E is invertible, suppose we have scalars $\alpha_1, \ldots, \alpha_n$ such that

$$\alpha_1 Eu_1 + \alpha_2 Eu_2 + \cdots + \alpha_n Eu_n = 0.$$

Applying T, we get

$$\alpha_{k+1} S u_{k+1} + \cdots + \alpha_n S u_n = 0,$$

where we used the fact that $TE = S$ and $Su_1 = \cdots = Su_k = 0$ because $u_1, \ldots, u_k \in \ker S$. Since $u_1, \ldots, u_n$ is a basis of V, the vectors $u_{k+1}, \ldots, u_n$ map to a spanning list of $\operatorname{im} S$. Since $\dim \operatorname{im} S = n - k$, it follows that $Su_{k+1}, \ldots, Su_n$ are linearly independent. Then from the previous equation, $\alpha_{k+1} = \cdots = \alpha_n = 0$. Substituting this back into the original equation, we get $\alpha_1 v_1 + \cdots + \alpha_k v_k = 0$, and since $v_1, \ldots, v_k$ are linearly independent, we must have $\alpha_1 = \cdots = \alpha_k = 0$. Hence, all the scalars are zero, proving that E is invertible.

Conversely, suppose there exists an invertible $E \in \operatorname{End}(V)$ such that $S = TE$. We have that $\operatorname{im} S \subseteq \operatorname{im} T$ by Exercise 3B26. Because E is invertible, we also have that $T = SE^{-1}$, so $\operatorname{im} T \subseteq \operatorname{im} S$ by the same reasoning. Hence, $\operatorname{im} S = \operatorname{im} T$. Similarly, if $w \in \operatorname{im} T$, then there exists some $v \in V$ such that $Tv = w$. Since E is invertible, there exists some $u \in V$ such that $Eu = v$. Then we have

$$Su = TEu = Tv = w,$$

and thus $w \in \operatorname{im} S$. Hence, $\operatorname{im} T \subseteq \operatorname{im} S$. Therefore, $\operatorname{im} S = \operatorname{im} T$. □

Solution 3D.8 Suppose there exist invertible $E_1 \in \operatorname{End}(V)$ and invertible $E_2 \in \operatorname{End}(W)$ such that $S = E_2 T E_1$. We claim that E_1 is an invertible map from $\ker S$ to $\ker T$. To see this, suppose $v, w \in \ker S$ and $E_1 v = E_1 w$. Then $v = w$ since E_1 is invertible, proving injectivity. To see surjectivity, let $u \in \ker T$. Since E_1 is invertible, there exists some $v \in V$ such that $E_1 v = u$. Then

$$Sv = E_2 T E_1 v = E_2 T u = E_2 0 = 0,$$

and thus $v \in \ker S$. Hence, E_1 is an invertible map from $\ker S$ to $\ker T$, and thus $\dim \ker S = \dim \ker T$.

Conversely, suppose $\dim \ker S = \dim \ker T$. Then we may construct an invertible linear map $E_1 \in \operatorname{End}(V)$ that maps $\ker S$ onto $\ker T$. We now claim that $\ker TE_1 = \ker S$. To see this, let $v \in \ker TE_1$. This is only true when $TE_1 v = 0$, or $E_1 v \in \ker T$. By invertibility of E_1, we have $v \in \ker S$. By Exercise 3D6 with $\ker S = \ker TE_1$, there exists an invertible $E_2 \in \operatorname{End}(W)$ such that $S = E_2 T E_1$. □

Solution 3D.9

Axler Note. *Here $T|_U$ means the function T restricted to U. Thus, $T|_U$ is the function whose domain is U, with $T|_U(u) = Tu$ for every $u \in U$.*

Since T is surjective, then $\operatorname{im} T = W$. By Exercise 3B11, there exists a subspace U of V such that $V = \ker T \oplus U$. We show that $T|_U$ is an isomorphism of U onto W.

Suppose $u \in U$ such that $T|_U u = 0$. Then $Tu = 0$, and thus $u \in \ker T \cap U$. Since $V = \ker T \oplus U$, we have $\ker T \cap U = \{0\}$, and thus $u = 0$. This shows that $T|_U$ is injective.

To see surjectivity, let $w \in W$. Since T is surjective, there exists $v \in V$ such that $Tv = w$. Write $v = n + u$ with $n \in \ker T$ and $u \in U$. Then $Tu = T(v - n) = Tv - Tn = w - 0 = w$. Hence, $T|_U$ is surjective onto W, and therefore an isomorphism of U onto W. □

Solution 3D.10

Axler Hint. *Define* $\Phi : \mathrm{Hom}(V, W) \to \mathrm{Hom}(U, W)$ *by* $\Phi(T) = T|_U$. *What is* $\ker \Phi$. *What is* $\mathrm{im}\, \Phi$*?*

(a) Consider the zero map $0 \in \mathrm{Hom}(V, W)$. For this map, $U \subseteq \ker 0$ since $\ker 0 = V$. Hence, $0 \in \mathscr{E}$, showing that $\mathscr{E}$ is non-empty. To see that $\mathscr{E}$ is closed under addition and scalar multiplication, let $S, T \in \mathscr{E}$ and $\alpha \in \mathbb{F}$. For any $u \in U$, we have $(S + T)u = Su + Tu = 0 + 0 = 0$, so $S + T \in \mathscr{E}$. Similarly, $(\alpha T)u = \alpha(Tu) = \alpha \cdot 0 = 0$, so $\alpha T \in \mathscr{E}$. Therefore, $\mathscr{E}$ is a subspace of $\mathrm{Hom}(V, W)$.

(b) Consider the linear map $\Phi : \mathrm{Hom}(V, W) \to \mathrm{Hom}(U, W)$ defined by $\Phi(T) = T|_U$. Let $T \in \mathrm{Hom}(V, W)$ such that $\Phi(T) = T|_U = 0$. Then $T|_U = 0$ implies that for every $u \in U$, we have $Tu = 0$, i.e., $U \subseteq \ker T$. Hence, $T \in \mathscr{E}$, showing that $\ker \Phi \subseteq \mathscr{E}$. Conversely, if $T \in \mathscr{E}$, then $U \subseteq \ker T$, which means that $T|_U = 0$, i.e., $T \in \ker \Phi$. Therefore, $\ker \Phi = \mathscr{E}$, and $\dim \mathscr{E} = \dim \ker \Phi$.

We now show that Φ is surjective. Let $S \in \mathrm{Hom}(U, W)$, and let $u_1, \ldots, u_m$ be a basis for U. Extend this to a basis $u_1, \ldots, u_m, v_1, \ldots, v_n$ for V. Define a linear map $T \in \mathrm{Hom}(V, W)$ by setting $Tu_i = Su_i$ for $i = 1, \ldots, m$ and $Tv_j = 0$ for $j = 1, \ldots, n$. Then $T|_U = S$, so $\Phi(T) = S$. Hence, Φ is surjective, and $\mathrm{im}\, \Phi = \mathrm{Hom}(U, W)$. Therefore, $\dim \mathrm{im}\, \Phi = \dim \mathrm{Hom}(U, W) = (\dim U)(\dim W)$. By the fundamental theorem of linear maps,

$$\dim \mathrm{Hom}(V, W) = \dim \ker \Phi + \dim \mathrm{im}\, \Phi = \dim \mathscr{E} + (\dim U)(\dim W),$$

Observing that $\dim \mathrm{Hom}(V, W) = (\dim V)(\dim W)$, we get

$$\dim \mathscr{E} = (\dim V)(\dim W) - (\dim U)(\dim W) = (\dim V - \dim U)(\dim W). \quad \square$$

Solution 3D.11 Suppose ST is invertible. Then $\ker ST = \{0\}$. If $v \in \ker T$, then $STv = S0 = 0$, and thus $v \in \ker ST$. Since $\ker ST = \{0\}$, we have $v = 0$, and thus $\ker T = \{0\}$. Hence, T is injective, and since V is finite-dimensional, T is invertible. Then $S = (ST)T^{-1}$ is a product of invertible linear maps, and thus S is invertible by Exercise 3D2.
Conversely, suppose S and T are invertible. Then ST is a product of invertible linear maps, and thus ST is invertible by Exercise 3D2. $\square$

Solution 3D.12 Since $STU = I$, then we have that $S(TU) = I$. By Exercise 3B20, then S is surjective, hence invertible with $S^{-1} = TU$. Similarly, $(ST)U = I$. By Exercise 3B19, then U is injective, hence invertible with $U^{-1} = ST$.

$$T(US) = (TU)S = S^{-1}S = I, \quad (US)T = U(ST) = UU^{-1} = I.$$

Hence, T is invertible with $T^{-1} = US$. $\square$

Solution 3D.13 If V were not finite-dimensional, let $v_1, v_2, \ldots$ be an infinite basis for V. Let $S, T, U \in \mathrm{End}(V)$ be defined as follows: let $S = I$,

$$Tv_1 = 0, \quad Tv_k = v_{k-1} \text{ for } k = 2, 3, \ldots$$

and

$$Uv_k = v_{k+1} \text{ for } k = 1, 2, \ldots$$

It is clear that $TUv_k = v_k$ for all $k = 1, 2, \ldots$, so $STU = I$. However, T is not injective since $\ker T = \operatorname{span} v_1$, and thus T is not invertible. □

Solution 3D.14 Since RST is surjective, and V is finite dimensional, it follows that RST is injective, hence invertible. By Exercise 3D11, then $RST = (RS)T$ is invertible if and only if RS and T are invertible. In particular, RS is invertible, and applying Exercise 3D11 again, we have that R and S are invertible. Hence, S is injective. □

Solution 3D.15 Since $Tv_1, \ldots, Tv_m$ spans V, we have T surjective. Moreover, V is finite-dimensional, so a surjective linear map on V is also injective, hence invertible. Therefore, for every $v \in V$, there exist scalars $\alpha_1, \ldots, \alpha_m$ such that

$$v = \alpha_1 Tv_1 + \cdots + \alpha_m Tv_m.$$

Applying T^{-1} to both sides, we get

$$T^{-1}v = \alpha_1 v_1 + \cdots + \alpha_m v_m.$$

Since T^{-1} is surjective, every vector $w \in V$ can be written as $T^{-1}v$ for some $v \in V$, and thus $v_1, \ldots, v_m$ spans V. □

Solution 3D.16 Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^{n,1}$. Define the k-th column of A to be $Te_k \in \mathbb{F}^{m,1}$ for $k = 1, \ldots, n$. Then for every $x \in \mathbb{F}^{n,1}$, there exist scalars $\alpha_1, \ldots, \alpha_n$ such that

$$x = \alpha_1 e_1 + \cdots + \alpha_n e_n.$$

Applying T to both sides, we have

$$Tx = \alpha_1 Te_1 + \cdots + \alpha_n Te_n = Ax,$$

where the last equality follows from the definition of matrix multiplication. Hence, $Tx = Ax$ for every $x \in \mathbb{F}^{n,1}$. □

Solution 3D.17

(a) Consider $T \in \ker \mathcal{A}$. Then $\mathcal{A}(T) = ST = 0$, so T is a linear map from V to $\ker S$. Conversely, if T is a linear map from V to $\ker S$, then $ST = 0$, so $T \in \ker \mathcal{A}$. Hence, $\ker \mathcal{A}$ is the set of linear maps from V to $\ker S$, and thus $\dim \ker \mathcal{A} = (\dim V)(\dim \ker S)$.

(b) By fundamental theorem of linear maps, we have

$$\dim \mathrm{End}(V) = \dim \ker \mathcal{A} + \dim \operatorname{im} \mathcal{A}.$$

Since $\dim \mathrm{End}(V) = (\dim V)^2$ and $\dim \ker \mathcal{A} = (\dim V)(\dim \ker S)$, we get

$$(\dim V)^2 = (\dim V)(\dim \ker S) + \dim \operatorname{im} \mathcal{A},$$

which implies

$$\dim \operatorname{im} \mathcal{A} = (\dim V)^2 - (\dim V)(\dim \ker S) = (\dim V)(\dim \operatorname{im} S). \qquad \square$$

Solution 3D.18 Let $T : \mathrm{Hom}(\mathbb{F}, V) \to V$ be defined by $T(\alpha) = \alpha(1)$ for $\alpha \in \mathrm{Hom}(\mathbb{F}, V)$. We first show that T is a linear map. Let $\alpha, \beta \in \mathrm{Hom}(\mathbb{F}, V)$ and $\lambda \in \mathbb{F}$. Then

$$T(\alpha + \beta) = (\alpha + \beta)(1) = \alpha(1) + \beta(1) = T\alpha + T\beta,$$

and

$$T(\lambda\alpha) = (\lambda\alpha)(1) = \lambda\alpha(1) = \lambda T\alpha.$$

We now show that T is an isomorphism. To see injectivity, let $\alpha, \beta \in \mathrm{Hom}(\mathbb{F}, V)$ such that $T\alpha = T\beta$. Then $\alpha(1) = \beta(1)$. For any $x \in \mathbb{F}$, we have $\alpha(x) = x\alpha(1) = x\beta(1) = \beta(x)$, so $\alpha = \beta$. Hence, T is injective.
To see surjectivity, let $v \in V$. Define $\alpha \in \mathrm{Hom}(\mathbb{F}, V)$ by $\alpha(x) = xv$ for $x \in \mathbb{F}$. Then $T\alpha = \alpha(1) = v$, so T is surjective. Therefore, T is an isomorphism, and thus V and $\mathrm{Hom}(\mathbb{F}, V)$ are isomorphic vector spaces. $\square$

Solution 3D.19

Appendix. *An alternate proof is given at the end of this solution.*

Suppose T has the same matrix with respect to every basis of V. Then for any set of bases for V, its matrix must be the same. If $\dim V \leq 1$, then every operator on V is a scalar multiple of the identity operator. Now suppose $\dim V \geq 2$. For any bases $v_1, \ldots, v_n$ and $u_1, \ldots, u_n$ of V, we have

$$\mathrm{Mat}(T, (v_1, \ldots, v_n)) = \mathrm{Mat}(T, (u_1, \ldots, u_n)).$$

Hence, we can declare that their matrix is just some A. We now consider the basis $v_1 + v_2, v_2, \ldots, v_n$, which yields the same matrix A. Let $w_1 = v_1 + v_2$ and $w_i = v_i$ for $i \geq 2$. In the new basis, we have the expression

$$Tw_1 = A_{1,1}w_1 + A_{2,1}w_2 + \cdots + A_{n,1}w_n = A_{1,1}v_1 + (A_{1,1} + A_{2,1})v_2 + A_{3,1}v_3 + \cdots + A_{n,1}v_n.$$

This must equal the expression for $Tw_1 = Tv_1 + Tv_2$ in the original basis:

$$Tv_1 + Tv_2 = (A_{1,1} + A_{1,2})v_1 + (A_{2,1} + A_{2,2})v_2 + \cdots + (A_{n,1} + A_{n,2})v_n.$$

Comparing coefficients of $v_1, \ldots, v_n$ in the two expressions, we get

$$\begin{aligned} A_{1,1} &= A_{1,1} + A_{1,2}, \\ A_{1,1} + A_{2,1} &= A_{2,1} + A_{2,2}, \\ A_{j,1} &= A_{j,1} + A_{j,2} \text{ for } j \geq 3. \end{aligned}$$

From this, we obtain that $A_{1,2} = 0, A_{2,2} = A_{1,1}$, and $A_{j,2} = 0$ for $j \geq 3$, i.e., the diagonal elements $A_{1,1}$ and $A_{2,2}$ are equal, while off-diagonal entries in column 2 are 0. Now consider the basis $v_1, v_1 + v_2, \ldots, v_n$. Let $u_1 = v_1, u_2 = v_1 + v_2$, and $u_i = v_i$ for $i \geq 3$. In the new basis, we have

$$\begin{aligned} Tu_2 &= A_{1,2}u_1 + \cdots + A_{n,2}u_n \\ &= A_{1,2}v_1 + A_{2,2}(v_1 + v_2) + A_{3,2}v_3 + \cdots + A_{n,2}v_n \\ &= (A_{1,2} + A_{2,2})v_1 + A_{2,2}v_2 + A_{3,2}v_3 + \cdots + A_{n,2}v_n. \\ &= A_{1,1}v_1 + A_{2,2}v_2. \end{aligned}$$

This must equal the expression for $Tu_2 = Tv_1 + Tv_2$ in the original basis, whose expression is listed above. Comparing coefficients of v_1 and v_2, we get

$$\begin{aligned} A_{1,1} &= A_{1,1} + A_{1,2}, \\ A_{2,2} &= A_{2,1} + A_{2,2}, \\ A_{j,2} &= A_{j,1} + A_{j,2} \quad \text{for } j \geq 3. \end{aligned}$$

From this, we obtain similar results about diagonal entries, but the off-diagonal entries in column 1 must also be 0. We may repeat the first process for $v_1 + v_k$ for $k \geq 3$ to get that entries off the diagonal in column k must also be 0. Continuing this process, we conclude all off-diagonal entries are 0, and all diagonal entries are equal, i.e., A is a scalar multiple of the identity matrix.

Conversely, suppose T is a scalar multiple of the identity operator. Then there exists some $\lambda \in \mathbb{F}$ such that $T = \lambda I$. For any basis of V, the matrix representation of T with respect to that basis is the diagonal matrix with λ on the diagonal. Hence, T has the same matrix with respect to every basis of V.

Alternate proof. The manual solution provides a constructive proof utilizing sums of the form $v_1 + v_k$ to eliminate the k-th column, along with $v_2 + v_1$ to eliminate the 1st column. Here, we present different constructions. Each proof assumes the declaration of the matrix to be A and that $\dim V \geq 2$.

This method shows that we can swap two *arbitrary* basis vectors to force the same value of all off-diagonal entries as well as the same value of all diagonal entries. Then, we can negate one basis vector to force all off-diagonal entries to be 0 via the linear independence of the basis vectors.

Alternate proof. Coolempire93. Let i, j be arbitrary with $1 \leq i, j \leq n$ and $i \neq j$. Consider the basis $w_1, \ldots, w_n$ of V defined by $w_i = v_j$, $w_j = v_i$ and $w_k = v_k$ otherwise. Then

$$\begin{aligned} T(w_j - v_i) &= Tw_j - Tv_i \\ T(0) &= A_{1,j}w_1 + A_{2,j}w_2 + \cdots + A_{n,j}w_n - (A_{1,i}v_1 + A_{2,i}v_2 + \cdots + A_{n,i}v_n) \\ 0 &= (A_{j,j} - A_{i,i})v_i + (A_{i,j} - A_{j,i})v_j + \sum_{\substack{k \neq i \\ k \neq j}} (A_{k,j} - A_{k,i})v_k. \end{aligned}$$

By linear independence of $v_1, \ldots, v_n$, we have $A_{i,i} = A_{j,j}$, $A_{i,j} = A_{j,i}$ and $A_{k,j} = A_{k,i}$ for all $k \neq i, j$. Because i and j were arbitrary, we have $A_{1,1} = A_{2,2} = \cdots = A_{n,n} = \lambda$, as well as A is symmetric, and all off-diagonal entries match across rows, so we can say

$A_{i,j} = \gamma$ for all $i \neq j$. To finish, consider the basis $u_1, \ldots, u_n$ defined by $u_1 = -v_1$ and $u_k = v_k$ otherwise. Then

$$\begin{aligned} T(v_1 + u_1) &= Tv_1 + Tu_1 \\ T(0) &= \lambda v_1 + \gamma(v_2 + \cdots + v_n) + (\lambda u_1 + \gamma(u_2 + \cdots + u_n)) \\ 0 &= 2\gamma(v_2 + \cdots + v_n). \end{aligned}$$

Again by linear independence of $v_1, \ldots, v_n$, we have $2\gamma = 0$ and thus $\gamma = 0$. Therefore, $A = \lambda I$ and T is a scalar multiple of the identity.

Alternate proof. This method shows that we can negate one *arbitrary* basis vector to force all off-diagonal entries to be 0 via the linear independence of the basis vectors. Then, we can swap v_1 with any other basis vector to force all diagonal entries to be the same.

Alternate proof. Coolempire93. Let j be arbitrary and consider the basis $w_1, \ldots, w_n$ of V defined by $w_j = -v_j$ and $w_i = v_i$ otherwise. Then

$$\begin{aligned} T(w_j + v_j) &= Tw_j + Tv_j \\ T(0) &= A_{1,j}w_1 + A_{2,j}w_2 + \cdots + A_{n,j}w_n + A_{1,j}v_1 + A_{2,j}v_2 + \cdots + A_{n,j}v_n \\ 0 &= 2\sum_{k \neq j} A_{k,j}v_k. \end{aligned}$$

By linear independence of $v_1, \ldots, v_n$, we have $A_{k,j} = 0$ for all $k \neq j$. Since j was arbitrary, we have that all off-diagonal entries of A are 0.

We turn to one more basis to prove that all diagonal entries of A must be the same. Once again, let j be arbitrary, and this time consider the basis $u_1, \ldots, u_n$ defined by $u_1 = v_j$, $u_j = v_1$ and $u_i = v_i$ otherwise. Then

$$\begin{aligned} T(v_1 - u_j) &= Tv_1 - Tu_j \\ T(0) &= A_{1,1}v_1 - A_{j,j}u_j \\ 0 &= (A_{1,1} - A_{j,j})v_1. \end{aligned}$$

As v_1 belongs to a basis for V, it is nonzero, and thus as j was arbitrary, $A_{1,1} = A_{2,2} = \cdots = A_{n,n} = \lambda$ for some scalar λ, and we have $A = \lambda I$, i.e., T is a multiple of the identity operator. □

Solution 3D.20

Appendix. *An alternate proof is given at the end of this solution.*

Let $T : \mathbb{R}[X]_{\leq n} \to \mathbb{R}[X]_{\leq n}$ be defined by

$$Tp = (x^2 + x)p'' + 2xp' + p(3)$$

for $p \in \mathbb{R}[X]_{\leq n}$. We first show that T is a linear map. Let $p, q \in \mathbb{R}[X]_{\leq n}$ and $\alpha \in \mathbb{R}$. Then

$$T(p + q) = (x^2 + x)(p + q)'' + 2x(p + q)' + (p + q)(3) = Tp + Tq,$$

and

$$T(\alpha p) = (x^2 + x)(\alpha p)'' + 2x(\alpha p)' + (\alpha p)(3) = \alpha Tp.$$

Then $T \in \text{End}(\mathbb{R}[X]_{\leq n})$. We now show that T is invertible. To see injectivity, let $p \in \mathbb{R}[X]_{\leq n}$ such that $Tp = 0$. Then

$$(x^2 + x)p'' + 2xp' + p(3) = 0.$$

Then

$$(x^2 + x)p'' + 2xp' = -p(3).$$

We analyze the degree of p as follows:

- If $\deg p \geq 2$, consider $(x^2 + x)p''$. In particular, $(a_m x^m)'' = a_m m(m-1)x^{m-2}$, hence the coefficient of x^m is $a_m m(m-1)$. Moreover, $2xp'$ has coefficient $2ma_m$. Since $\deg p \geq 2$, then $m \geq 2$, so the total coefficient of x^m is $a_m m(m-1) + 2ma_m = a_m m(m-1+2) = a_m m(m+1)$. Since this is nonzero for $a_m \neq 0$, we get a contradiction with the right-hand side being a constant.
- If $\deg p = 1$, then $p(x) = ax + b$ for some $a, b \in \mathbb{R}$. Then $p''(x) = 0$ and $p'(x) = a$, so the left-hand side is $2ax$. Since the right-hand side is a constant, this is a contradiction.
- If $\deg p = 0$, then $p(x) = c$ for some $c \in \mathbb{R}$. Then $p''(x) = 0$ and $p'(x) = 0$, so the left-hand side is 0, and the right-hand side is $p(3) = c$. Hence, $c = 0$, and thus $p = 0$. Therefore, T is injective.

Since T is injective and $\mathbb{R}[X]_{\leq n}$ is finite-dimensional, then T is surjective, hence invertible. Then there does exist some $p \in \mathbb{R}[X]_{\leq n}$ such that $Tp = q$.

Alternate proof. Coolempire93. We go by the same technique as in the alternate proof of Exercise 3B29.

Let $q(x) = a_0 + a_1 x + \cdots + a_n x^n$ and suppose that there exists a polynomial $p(x) = b_0 + b_1 x + \cdots + b_m x^m$ satisfying the constraint $q(x) = (x^2+x)p''(x) + 2xp'(x) + p(3)$. If q is constant, then let $p = q = a_0$. We have $(x^2 + x)p''(x) + 2xp'(x) + p(3) = (x^2 + x)0 + 2x(0) + a_0 = a_0 = q(x)$. If q is linear, say $q(x) = a_0 + a_1 x$, then let $p(x) = b_0 + b_1 x$. We want

$$q(x) = (x^2+x)p''(x) + 2xp'(x) + p(3) = (x^2+x)0 + 2x(b_1) + b_0 + 3b_1 = 2b_1 x + 3b_1 + b_0,$$

so letting $b_1 = \frac{a_1}{2}$ and $b_0 = a_0 - \frac{3a_1}{2}$ yields a polynomial p satisfying the given equation. Otherwise, consider when $n \geq 2$. If p has degree $m \geq 2$, then p' has degree $m-1$ and p'' has degree $m-2$, so $(x^2 + x)p'' + 2xp'$ has degree m. To satisfy the constraint, we must have $m = n \geq 2$, so we may write $p(x) = b_0 + b_1 x + \cdots + b_n x^n$. Evaluating, we obtain

$$p'(x) = b_1 + 2b_2 x + \cdots + nb_n x^{n-1} \implies 2xp'(x) = 2b_1 x + 4b_2 x^2 + \cdots + 2nb_n x^n$$

$$p''(x) = 2b_2 + 6b_3 x + \cdots + n(n-1)b_n x^{n-2} \implies \begin{cases} x^2 p''(x) = 2b_2 x^2 + 6b_3 x^3 + \cdots + n(n-1)b_n x^n \\ xp''(x) = 2b_2 x + 6b_3 x^2 + \cdots + n(n-1)b_n x^{n-1} \end{cases}.$$

Equating coefficients of $(x^2 + x)p'' + 2xp' + p(3)$ with those of q, we have that

$$\begin{cases} a_0 = p(3) = b_0 + 3b_1 + 9b_2 + \cdots + 3^n b_n \\ a_i = i(i+1)(b_i + b_{i+1}) \qquad \text{for } 1 \leq i \leq n-1 \\ a_n = n(n-1)b_n + 2nb_n = n(n+1)b_n \end{cases},$$

with the formula for a_i coming from the fact that the coefficient of x^i in $(x^2+x)p''(x)+2xp'(x)$ is

$$(i(i-1)b_i + i(i+1)b_{i+1}) + 2ib_i = (i^2 - i + 2i)b_i + i(i+1)b_{i+1} = i(i+1)(b_i + b_{i+1}).$$

As before, we solve for the coefficients of p by back-substitution. Noting that for $1 \le i \le n$, we have

$$b_i = \frac{a_i}{i(i+1)} - b_{i+1},$$

we obtain the system

$$\begin{cases} b_n = \frac{a_n}{n(n+1)} \\ b_{n-1} = \frac{a_{n-1}}{n(n-1)} - b_n = \frac{a_{n-1}}{n(n-1)} - \frac{a_n}{n(n+1)} \\ b_{n-2} = \frac{a_{n-2}}{(n-1)(n-2)} - b_{n-1} = \frac{a_{n-2}}{(n-1)(n-2)} - \frac{a_{n-1}}{n(n-1)} + \frac{a_n}{n(n+1)} \\ \vdots \\ b_1 = \frac{a_1}{2} - b_2 \\ b_0 = a_0 - 3b_1 - 9b_2 - \cdots - 3^n b_n \end{cases}$$

which has a valid solution† since no division by 0 occurs.

† As before, solving this recurrence is outside the scope of this course, but, specifically, the coefficients of p are

$$b_i = \frac{a_i}{i(i+1)} - \frac{a_{i+1}}{(i+1)(i+2)} + \cdots + (-1)^{n-i}\frac{a_n}{n(n+1)} = \sum_{j=i}^{n}(-1)^{j-i}\frac{a_j}{j(j+1)}$$

for $1 \le i \le n$ and $b_0 = a_0 - \sum_{j=1}^{n} 3^j b_j$. □

Solution 3D.21

(a) Suppose the trivial solution $0 \in \mathbb{F}^{n,1}$ is the only solution to the homogeneous system of equations. Then $\ker T = \{0\}$, where $T : \mathbb{F}^{n,1} \to \mathbb{F}^{n,1}$ is defined by $Tx = Ax$ for $x \in \mathbb{F}^{n,1}$. Since $\ker T = \{0\}$, then T is injective. Since $\mathbb{F}^{n,1}$ is finite-dimensional, then T is surjective, hence invertible. Then for every $c_1, \ldots, c_n \in \mathbb{F}$, there exists some $x \in \mathbb{F}^{n,1}$ such that $Tx = c$, where c is the column vector with entries $c_1, \ldots, c_n$. Hence, there exists a solution to the system of equations.

(b) Suppose for every $c_1, \ldots, c_n \in \mathbb{F}$, there exists a solution to the system of equations. Then $T : \mathbb{F}^{n,1} \to \mathbb{F}^{n,1}$ defined by $Tx = Ax$ for $x \in \mathbb{F}^{n,1}$ is surjective. Since $\mathbb{F}^{n,1}$ is finite-dimensional, then T is injective. Hence, the trivial solution is the only solution to the homogeneous system of equations.

□

Solution 3D.22 Suppose $\mathrm{Mat}(T, (v_1, \ldots, v_n))$ is invertible. Then there exists some $A \in \mathbb{F}^{n,n}$ such that $A\,\mathrm{Mat}(T, (v_1, \ldots, v_n)) = I$. Let U be the linear map from V to V whose matrix with respect to the basis $v_1, \ldots, v_n$ is A. Then

$$\mathrm{Mat}(UT, (v_1, \ldots, v_n)) = \mathrm{Mat}(U, (v_1, \ldots, v_n))\,\mathrm{Mat}(T, (v_1, \ldots, v_n)) = A\,\mathrm{Mat}(T, (v_1, \ldots, v_n)) = I.$$

Then UT is the identity operator, and thus T is injective. Since V is finite-dimensional, then T is surjective, hence invertible.
Suppose T is invertible. Then T^{-1} exists, and we have

$$\operatorname{Mat}(T^{-1}T, (v_1, \dots, v_n)) = \operatorname{Mat}(T^{-1}, (v_1, \dots, v_n)) \operatorname{Mat}(T, (v_1, \dots, v_n)) = I.$$

Hence, $\operatorname{Mat}(T, (v_1, \dots, v_n))$ is invertible. □

Solution 3D.23 By definition, column k of $\operatorname{Mat}(T, (v_1, \dots, v_n))$ is $Tv_k = u_k$ with respect to the basis $v_1, \dots, v_n$. Similarly, column k of $\operatorname{Mat}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$ is also u_k with respect to the basis $v_1, \dots, v_n$. □

Solution 3D.24 Let $T : \mathbb{F}^{n,1} \to \mathbb{F}^{n,1}$ be defined by $Tx = Ax$ for $x \in \mathbb{F}^{n,1}$. Since $AB = I$, then T is surjective. Since $\mathbb{F}^{n,1}$ is finite-dimensional, surjectivity implies injectivity. Hence, T is invertible, and $T(Bx) = x = T(T^{-1}x)$. Then $Bx = T^{-1}x$ for every $x \in \mathbb{F}^{n,1}$. Therefore,

$$BAx = B(Tx) = T^{-1}(Tx) = x$$

for every $x \in \mathbb{F}^{n,1}$. Hence, $BA = I$. □

§3E

Solution 3E.1 Let $G(T)$ denote the graph of T. Now, suppose T is a linear map. Then $G(T)$ is nonempty since $(0, T0) = (0, 0) \in G(T)$. Now, let $(v_1, Tv_1), (v_2, Tv_2) \in G(T)$. Then

$$(v_1 - v_2, Tv_1 - Tv_2) = (v_1, Tv_1) - (v_2, Tv_2) \in G(T)$$

since T is linear. Now, let $(v, Tv) \in G(T)$ and $\alpha \in \mathbb{F}$. Then

$$(\alpha v, \alpha Tv) = \alpha(v, Tv) \in G(T)$$

since T is linear. Thus, $G(T)$ is a subspace of $V \times W$.
Conversely, suppose $G(T)$ is a subspace of $V \times W$. By definition, we have $(v_1, Tv_1), (v_2, Tv_2) \in G(T)$ for all $v_1, v_2 \in V$. Since $G(T)$ is a subspace of $V \times W$, for all $\alpha, \beta \in \mathbb{F}$, we have

$$(\alpha v_1 + \beta v_2, \alpha Tv_1 + \beta Tv_2) = \alpha(v_1, Tv_1) + \beta(v_2, Tv_2) \in G(T).$$

By definition of $G(T)$, we have $\alpha Tv_1 + \beta Tv_2 = T(\alpha v_1 + \beta v_2)$. Thus, T is a linear map. □

Solution 3E.2 Let $V = V_1 \times \cdots \times V_m$, and consider $T \in \operatorname{Hom}(V, V_k)$ defined by $T(v_1, \dots, v_m) = v_k$. Then T is a linear map since for all $\alpha, \beta \in \mathbb{F}$ and all $(v_1, \dots, v_m), (w_1, \dots, w_m) \in V$, we have

$$\begin{aligned} T(\alpha(v_1, \dots, v_m) + \beta(w_1, \dots, w_m)) &= T(\alpha v_1 + \beta w_1, \dots, \alpha v_m + \beta w_m) \\ &= \alpha v_k + \beta w_k \\ &= \alpha T(v_1, \dots, v_m) + \beta T(w_1, \dots, w_m). \end{aligned}$$

Moreover, T is surjective since for any $v_k \in V_k$, we have $v_k = T(0, \ldots, 0, v_k, 0, \ldots, 0)$ where v_k is in the k-th coordinate. Since V is finite-dimensional and T is a linear map that is surjective, it follows that V_k is finite-dimensional. □

Solution 3E.3 Let $V = V_1 \times \cdots \times V_m$. Let $L = \text{Hom}(V_1, W) \times \cdots \times \text{Hom}(V_m, W)$. Define $T \in \text{Hom}(\text{Hom}(V, W), L)$ by

$$T(S) = (S \circ \iota_1, \ldots, S \circ \iota_m)$$

where $\iota_k \in \text{Hom}(V_k, V)$ is the map defined as

$$\iota_k(v_k) = (0, \ldots, 0, v_k, 0, \ldots, 0)$$

with $\iota_k(v_k)$ having v_k in the k-th coordinate and 0 in all other coordinates. Then T is linear since for all $\alpha, \beta \in \mathbb{F}$ and all $S_1, S_2 \in \text{Hom}(V, W)$, we have

$$\begin{aligned} T(\alpha S_1 + \beta S_2) &= ((\alpha S_1 + \beta S_2) \circ \iota_1, \ldots, (\alpha S_1 + \beta S_2) \circ \iota_m) \\ &= (\alpha(S_1 \circ \iota_1) + \beta(S_2 \circ \iota_1), \ldots, \alpha(S_1 \circ \iota_m) + \beta(S_2 \circ \iota_m)) \\ &= \alpha T(S_1) + \beta T(S_2), \end{aligned}$$

hence T is a linear map. Now suppose $T(S) = 0$. Then $S \circ \iota_k = 0$ for all $k = 1, \ldots, m$. Moreover, for all $(v_1, \ldots, v_m) \in V$, we have

$$S(v_1, \ldots, v_m) = S(\iota_1(v_1) + \cdots + \iota_m(v_m)) = S \circ \iota_1(v_1) + \cdots + S \circ \iota_m(v_m) = 0.$$

It follows that $S(v_1, \ldots, v_m) = 0$ for all $(v_1, \ldots, v_m) \in V$. Thus, $S = 0$, and hence T is injective.
Now, let $(S_1, \ldots, S_m) \in L$. Define $S \in \text{Hom}(V, W)$ by $S(v_1, \ldots, v_m) = S_1 v_1 + \cdots + S_m v_m$. Then $T(S) = (S_1, \ldots, S_m)$. Thus, T is surjective. Therefore, T is an isomorphism. □

Solution 3E.4

Manual Remark. *Note the difference between this exercise and Exercise 3E3. Here, the domain is decomposed, while in Exercise 3E3, the codomain is decomposed. Hence, the use of the maps ι_k and π_k in the solutions to these exercises are reversed.*

Let $W = W_1 \times \cdots \times W_m$. Let $L = \text{Hom}(V, W_1) \times \cdots \times \text{Hom}(V, W_m)$. Define $T \in \text{Hom}(\text{Hom}(V, W), L)$ by

$$T(S) = (\pi_1 \circ S, \ldots, \pi_m \circ S)$$

where $\pi_k \in \text{Hom}(W, W_k)$ is the map defined as

$$\pi_k(w_1, \ldots, w_m) = w_k$$

Then T is linear since for all $\alpha, \beta \in \mathbb{F}$ and all $S_1, S_2 \in \text{Hom}(V, W)$, we have

$$\begin{aligned} T(\alpha S_1 + \beta S_2) &= (\pi_1 \circ (\alpha S_1 + \beta S_2), \ldots, \pi_m \circ (\alpha S_1 + \beta S_2)) \\ &= (\alpha(\pi_1 \circ S_1) + \beta(\pi_1 \circ S_2), \ldots, \alpha(\pi_m \circ S_1) + \beta(\pi_m \circ S_2)) \\ &= \alpha T(S_1) + \beta T(S_2), \end{aligned}$$

hence T is a linear map. Now suppose $T(S) = 0$. Then $\pi_k \circ S = 0$ for all $k = 1, \ldots, m$. Moreover, for all $v \in V$, we have

$$Sv = (\pi_1(Sv), \ldots, \pi_m(Sv)) = (0, \ldots, 0) = 0.$$

It follows that $Sv = 0$ for all v, hence $S = 0$, and thus T is injective.

Now let $(S_1, \ldots, S_m) \in L$. Define $S \in \text{Hom}(V, W)$ by $Sv = (S_1 v, \ldots, S_m v)$. Then $T(S) = (S_1, \ldots, S_m)$. Thus, T is surjective. Therefore, T is an isomorphism. □

Solution 3E.5 Let $e_1, \ldots, e_m$ be the standard basis of $\mathbb{F}^m$. Define $T \in \text{Hom}(V^m, \text{Hom}(\mathbb{F}^m, V))$ by

$$T(v_1, \ldots, v_m)(\lambda_1, \ldots, \lambda_m) = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

Then T is linear since for all $\alpha, \beta \in \mathbb{F}$ and all $v = (v_1, \ldots, v_m), w = (w_1, \ldots, w_m) \in V^m$, we have

$$\begin{aligned} T(\alpha v + \beta w)(\lambda_1, \ldots, \lambda_m) &= T(\alpha v_1 + \beta w_1, \ldots, \alpha v_m + \beta w_m)(\lambda_1, \ldots, \lambda_m) \\ &= \lambda_1(\alpha v_1 + \beta w_1) + \cdots + \lambda_m(\alpha v_m + \beta w_m) \\ &= \alpha(\lambda_1 v_1 + \cdots + \lambda_m v_m) + \beta(\lambda_1 w_1 + \cdots + \lambda_m w_m) \\ &= \alpha T(v)(\lambda_1, \ldots, \lambda_m) + \beta T(w)(\lambda_1, \ldots, \lambda_m), \end{aligned}$$

hence T is a linear map. Now suppose $T(v) = 0$. Then $T(v)(e_k) = 0$ for all $k = 1, \ldots, m$. It follows that $v_k = 0$ for all $k = 1, \ldots, m$, hence $v = 0$, and thus T is injective.

Now let $S \in \text{Hom}(\mathbb{F}^m, V)$. Define $v \in V^m$ by $v_k = Se_k$ for all $k = 1, \ldots, m$. Then $T(v) = S$. Thus, T is surjective. Therefore, T is an isomorphism. □

Solution 3E.6 Let $u \in U$. Then $v + u \in v + U = x + W$. Thus, there exists $w \in W$ such that $v + u = x + w$. Note that $v \in x + W$ implies the existence of some $w' \in W$ such that $v = x + w'$, or $x - v = -w' \in W$. Similarly, $x \in v + U$ implies the existence of some $u' \in U$ such that $x = v + u'$, or $v - x = -u' \in U$. It follows that $u = (x - v) + w \in W$, hence $U \subseteq W$. Now let $w \in W$. Then $x + w \in x + W = v + U$. Thus, there exists $u \in U$ such that $x + w = v + u$. It follows that $w = (v - x) + u \in U$, hence $W \subseteq U$. Therefore, $U = W$. □

Solution 3E.7 Suppose A is a translate of U. Then there exists $v \in \mathbb{R}^3$ such that $A = v + U$. Let $v = (x_0, y_0, z_0)$. Then

$$\begin{aligned} A &= v + U \\ &= \{(x_0, y_0, z_0) + (x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = 0\} \\ &= \{(x', y', z') \in \mathbb{R}^3 : 2(x' - x_0) + 3(y' - y_0) + 5(z' - z_0) = 0\} \\ &= \{(x', y', z') \in \mathbb{R}^3 : 2x' + 3y' + 5z' = 2x_0 + 3y_0 + 5z_0\}. \end{aligned}$$

Conversely, suppose there exists $c \in \mathbb{R}$ such that

$$A = \{(x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = c\}.$$

Let $v = (x_0, y_0, z_0)$ where x_0, y_0, z_0 are any real numbers such that $2x_0+3y_0+5z_0 = c$. Then

$$\begin{aligned} v + U &= (x_0, y_0, z_0) + U \\ &= \{(x_0, y_0, z_0) + (x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = 0\} \\ &= A. \end{aligned}$$

□

Solution 3E.8

(a) Suppose $T \in \text{Hom}(V, W)$ and $c \in W$. Let $A = \{x \in V : Tx = c\}$. If A is nonempty, then there exists $v \in V$ such that $Tv = c$. We claim that $A = v + \ker T$. Let $x \in A$. Then $Tx = c = Tv$, hence $T(x - v) = 0$, and thus $x - v \in \ker T$. It follows that $x \in v + \ker T$, hence $A \subseteq v + \ker T$. Now let $y \in v + \ker T$. Then there exists $u \in \ker T$ such that $y = v + u$. It follows that $Ty = T(v + u) = Tv + Tu = c + 0 = c$, hence $y \in A$, and thus $v + \ker T \subseteq A$. Therefore, $A = v + \ker T$.

(b) A system of linear equations such as 3.27 can be written in the form $Tx = c$ for some linear map T and some vector c. By part (a), the set of solutions to this system of linear equations is either the empty set or is a translate of $\ker T$, which is a subspace of $\mathbb{F}^n$.

□

Solution 3E.9 Suppose A is a translate of some subspace of V. Then there exists $v \in V$ and a subspace U of V such that $A = v + U$. Let $v + u, v + u' \in A$. Then for any $\lambda \in \mathbb{F}$,

$$\lambda(v + u) + (1 - \lambda)(v + u') = v + \underbrace{(\lambda u + (1 - \lambda)u')}_{\in U} \in A.$$

Conversely, suppose $\lambda v + (1 - \lambda)w \in A$ for all $v, w \in A$ and all $\lambda \in \mathbb{F}$. Fix some $v_0 \in A$, and let $U = \{u \in V : v_0 + u \in A\}$. We first claim that U is a subspace of V. Since $v_0 = v_0 + 0 \in A$, we have $0 \in U$. Let $u_1, u_2 \in U$. Then $v_0 + u_1, v_0 + u_2 \in A$. Now let $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} \lambda(v_0 + u_1) + (1 - \lambda)(v_0 + u_2) &= v_0 + \lambda u_1 + (1 - \lambda)u_2 \\ &= v_0 + (\lambda u_1 + (1 - \lambda)u_2) \in A. \end{aligned}$$

It follows that $\lambda u_1 + (1 - \lambda)u_2 \in U$, hence U is a subspace of V. Now we claim that $A = v_0 + U$. Let $w \in A$. By definition, $v_0 + (w - v_0) = w \in A$, hence $w - v_0 \in U$, and thus $w \in v_0 + U$, so $A \subseteq v_0 + U$. Now let $v_0 + u \in v_0 + U$. Then $v_0 + u \in A$ by definition of U, hence $v_0 + U \subseteq A$. Therefore, $A = v_0 + U$. □

Solution 3E.10 If $A_1 \cap A_2$ is the empty set, then we are done. Suppose that $A_1 \cap A_2$ is nonempty. Let $x, y \in A_1 \cap A_2$ and $\lambda \in \mathbb{F}$. Then $x, y \in A_1$, hence $\lambda x + (1 - \lambda)y \in A_1$. Moreover, $x, y \in A_2$, hence $\lambda x + (1-\lambda)y \in A_2$. It follows that $\lambda x + (1-\lambda)y \in A_1 \cap A_2$. Thus, by Exercise 3E9, $A_1 \cap A_2$ is a translate of some subspace of V. □

Solution 3E.11

Appendix. *An alternate proof is given at the end of this solution.*

(a) We claim that U can be reformulated as the set of vectors $(x_1, x_2, \ldots) \in \mathbb{F}^\infty$ such that there exists an index N such that $x_k = 0$ for all $k > N$. To see this, suppose $(x_1, x_2, \ldots) \in U$. Then there exists a finite set of indices K such that $x_k = 0$ for all $k \notin K$. If K is empty, let $N = 0$. Otherwise, let $N = \max K$. Then $x_k = 0$ for all $k > N$.

Conversely, suppose there exists an index N such that $x_k = 0$ for all $k > N$. Then the set of indices $\{1, 2, \ldots, N\}$ is finite and contains all indices where $x_k \neq 0$, hence $(x_1, x_2, \ldots) \in U$. Thus, the two formulations of U are equivalent.

We now show that U is a subspace of $\mathbb{F}^\infty$. First, note that the zero vector $(0, 0, \ldots) \in U$ since it has no nonzero coordinates. Now let $u = (u_1, u_2, \ldots), v = (v_1, v_2, \ldots) \in U$. Then there exist indices N_u and N_v such that $u_k = 0$ for all $k > N_u$ and $v_k = 0$ for all $k > N_v$. Let $N = \max(N_u, N_v)$. Then for all $k > N$, we have $u_k = 0$ and $v_k = 0$, hence $(u + v)_k = u_k + v_k = 0 + 0 = 0$. Thus, $u + v \in U$. Now let $\alpha \in \mathbb{F}$. Then for all $k > N_u$, we have $(\alpha u)_k = \alpha u_k = \alpha \cdot 0 = 0$, hence $\alpha u \in U$. Therefore, U is a subspace of $\mathbb{F}^\infty$.

(b) For any given positive integer m, define the sets B_i for $1 \leq i \leq m$ as follows:

$$B_i = \{i + km : k \in \mathbb{N}\} = \{i, i + m, i + 2m, \ldots\}.$$

Then $B_1, \ldots, B_m$ are pairwise disjoint subsets of $\mathbb{N}$. For each $i = 1, \ldots, m$, define

$$v_i = (v_{i,1}, v_{i,2}, \ldots) \in \mathbb{F}^\infty$$

such that $v_{i,j} = 1$ if $j \in B_i$ and $v_{i,j} = 0$ otherwise. We claim that $v_1 + U, \ldots, v_m + U$ are linearly independent in $\mathbb{F}^\infty / U$.

Let $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ such that $\lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U) = U$. Then $\lambda_1 v_1 + \cdots + \lambda_m v_m \in U$. Then we can show that $\lambda_i = 0$ for all $i = 1, \ldots, m$ as follows. For each i, consider the j-th coordinate of v_i for $j \in B_i$. Since v_i has 1's in all coordinates indexed by B_i and 0's elsewhere, and B_i is disjoint from B_k for $k \neq i$, the j-th coordinate of $\lambda_1 v_1 + \cdots + \lambda_m v_m$ is λ_i. But this sum is in U, which means it has only finitely many nonzero coordinates. Since B_i is infinite, it follows that $\lambda_i = 0$. Hence, all $\lambda_i = 0$, proving that $v_1 + U, \ldots, v_m + U$ are linearly independent. Since m was arbitrary, $\mathbb{F}^\infty / U$ is infinite-dimensional.

Alternate proof. Civil Service Pigeon. This solution assumes that the reader is familiar with partitioning $\mathbb{N}$ into infinitely many disjoint infinite sets.

(a) Since $(0, 0, \ldots) \in U$, U is nonempty. Suppose $x, y \in U$ and $c \in \mathbb{F}$. Let S_x and S_y be the sets of indices where x and y have nonzero terms, respectively. By definition, S_x and S_y are finite. $cx + y$ can only have nonzero terms at indices in $S_x \cup S_y$. Since the union of two finite sets is finite, $cx + y \in U$, which makes U a subspace of $\mathbb{F}^\infty$.

(b) Partition $\mathbb{Z}^+$ into a countably infinite family of disjoint infinite subsets A_1, A_2, ... For each n, define $v_n \in \mathbb{F}^\infty$ such that its j-th term is 1 if $j \in A_n$ and 0 otherwise.

Suppose that $\sum_{i=1}^{m} c_i(v_i + U) = 0 + U$ for some positive integer m and scalars $c_1, \ldots, c_m \in \mathbb{F}$. Then, $w = \sum_{i=1}^{m} c_i v_i \in U$. By construction, for $i \in \{1, \ldots, m\}$, the j-th term of w is c_i for all $j \in A_i$. Since $w \in U$, it only has a finite number of nonzero terms. But each A_i is infinite, so this forces $c_i = 0$ for all $i \in \{1, \ldots, m\}$. Hence, $v_1 + U$, $v_2 + U$, ... are linearly independent in $\mathbb{F}^\infty/U$.

Because $\mathbb{F}^\infty/U$ contains an infinite sequence of linearly independent vectors, it is infinite-dimensional. □

Solution 3E.12

(a) For any element $\lambda_1 v_1 + \cdots + \lambda_m v_m \in A$, we can write

$$\lambda_1 v_1 + \cdots + \lambda_m v_m = v_1 + \lambda_2(v_2 - v_1) + \cdots + \lambda_m(v_m - v_1) = v_1 + \sum_{i=2}^{m} \lambda_i(v_i - v_1),$$

where $\lambda_1 = 1 - (\lambda_2 + \cdots + \lambda_m)$. Let $U = \operatorname{span}(v_2 - v_1, \ldots, v_m - v_1)$. Then $A = v_1 + U$, hence A is a translate of the subspace U of V.

(b) By Exercise 3E9, since B is a translate of some subspace of V, we have $\lambda v + (1 - \lambda)w \in B$ for all $v, w \in B$ and all $\lambda \in \mathbb{F}$. Since $\{v_1, \ldots, v_m\} \subseteq B$, it follows that any linear combination of $v_1, \ldots, v_m$ with coefficients summing to 1 is in B. Hence, $A \subseteq B$.

(c) The subspace $U = \operatorname{span}(v_2 - v_1, \ldots, v_m - v_1)$ has dimension at most $m - 1$, since it is spanned by $m - 1$ vectors. Therefore, $A = v_1 + U$ is a translate of a subspace of V of dimension less than m.

□

Solution 3E.13 Since V/U is finite-dimensional, let $v_1 + U, \ldots, v_m + U$ be a basis of V/U. Let $W = \operatorname{span}(v_1, \ldots, v_m)$. We show that $V = U \oplus W$. First, note that any $v \in V$ can be written as $v = u + w$ for some $u \in U$ and $w \in W$. Indeed, since $v + U \in V/U$, we can write

$$v + U = \lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U)$$

for some scalars $\lambda_1, \ldots, \lambda_m$. This gives

$$v - (\lambda_1 v_1 + \cdots + \lambda_m v_m) \in U.$$

So setting $u = v - (\lambda_1 v_1 + \cdots + \lambda_m v_m) \in U$ and $w = \lambda_1 v_1 + \cdots + \lambda_m v_m \in W$, we have $v = u + w$. To see that the sum is direct, suppose $u + w = 0$ with $u \in U$ and $w \in W$. Then $w = -u \in W \cap U$. Since $w \in W$, we can write $w = \lambda_1 v_1 + \cdots + \lambda_m v_m$ for some scalars $\lambda_1, \ldots, \lambda_m$. Since we established that $w \in U$, we have $w + U = 0 + U$. But $v_1 + U, \ldots, v_m + U$ are linearly independent in V/U, hence $\lambda_1 = \cdots = \lambda_m = 0$, and thus $w = 0$. It follows that $u = 0$ as well, proving that the sum is direct. Therefore, $V = U \oplus W$.

Consider the map $T : V \to U \times (V/U)$ given by $Tv = (u, v + U)$, where $v = u + w$ with $u \in U$ and $w \in W$. This map is well-defined because the decomposition $v = u + w$ is unique. It is linear, injective, and surjective, hence an isomorphism. Therefore, $V \cong U \times (V/U)$. □

Solution 3E.14 Since $V = U \oplus W$, every $v \in V$ can be written uniquely as $v = u + w$ with $u \in U$ and $w \in W$. We claim that $w_1 + U, \ldots, w_m + U$ is a basis of V/U. First, we show that they span V/U. Let $v + U \in V/U$. Write $v = u + w$ with $u \in U$ and $w \in W$. Then $v + U = (u + w) + U = w + U$. Since $w \in W$, we can write $w = \lambda_1 w_1 + \cdots + \lambda_m w_m$ for some scalars $\lambda_1, \ldots, \lambda_m$. Hence, $v + U = \lambda_1(w_1 + U) + \cdots + \lambda_m(w_m + U)$, showing that $w_1 + U, \ldots, w_m + U$ span V/U.
Next, we show linear independence. Suppose $\lambda_1(w_1 + U) + \cdots + \lambda_m(w_m + U) = 0 + U$. Then $\lambda_1 w_1 + \cdots + \lambda_m w_m \in U$. But $\lambda_1 w_1 + \cdots + \lambda_m w_m \in W$ and $U \cap W = \{0\}$, so $\lambda_1 w_1 + \cdots + \lambda_m w_m = 0$. Since $w_1, \ldots, w_m$ are linearly independent, we have $\lambda_1 = \cdots = \lambda_m = 0$. Therefore, $w_1 + U, \ldots, w_m + U$ are linearly independent, and hence form a basis of V/U. □

Solution 3E.15 First, we show that $v_1, \ldots, v_m, u_1, \ldots, u_n$ span V. Let $v \in V$. Then $v + U \in V/U$, so we can write

$$v + U = \lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U)$$

for some scalars $\lambda_1, \ldots, \lambda_m$. This gives $v - (\lambda_1 v_1 + \cdots + \lambda_m v_m) \in U$. Since $u_1, \ldots, u_n$ is a basis of U, we can write $v - (\lambda_1 v_1 + \cdots + \lambda_m v_m) = \mu_1 u_1 + \cdots + \mu_n u_n$ for some scalars $\mu_1, \ldots, \mu_n$. Hence, $v = \lambda_1 v_1 + \cdots + \lambda_m v_m + \mu_1 u_1 + \cdots + \mu_n u_n$, showing that $v_1, \ldots, v_m, u_1, \ldots, u_n$ span V.
Next, we show linear independence. Suppose $\lambda_1 v_1 + \cdots + \lambda_m v_m + \mu_1 u_1 + \cdots + \mu_n u_n = 0$. Then $\lambda_1 v_1 + \cdots + \lambda_m v_m \in U$. Then

$$\lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U) = 0 + U.$$

Since $v_1 + U, \ldots, v_m + U$ are linearly independent in V/U, we must have $\lambda_1 = \cdots = \lambda_m = 0$. Then the original equation reduces to $\mu_1 u_1 + \cdots + \mu_n u_n = 0$, and since $u_1, \ldots, u_n$ are linearly independent, we have $\mu_1 = \cdots = \mu_n = 0$. Therefore, $v_1, \ldots, v_m, u_1, \ldots, u_n$ are linearly independent, and hence form a basis of V. □

Solution 3E.16 From 3.107 in the text, we know that $V/\ker\phi \cong \operatorname{im}\phi$. Since $\phi \neq 0$, we have $\operatorname{im}\phi \neq \{0\}$, and since $\operatorname{im}\phi$ is a subspace of $\mathbb{F}$, it follows that $\dim\operatorname{im}\phi = 1$. This implies $\operatorname{im}\phi = \mathbb{F}$, and hence $\dim V/(\ker\phi) = \dim\operatorname{im}\phi = 1$. □

Solution 3E.17 Since $\dim V/U = \dim\mathbb{F} = 1$, there exists some isomorphism $\psi \in \operatorname{Hom}(V/U, \mathbb{F})$. Let $\pi : V \to V/U$ be the quotient map. Define $\phi = \psi \circ \pi \in \operatorname{Hom}(V, \mathbb{F})$. Suppose $\phi(v) = 0$. Then $\psi(\pi(v)) = 0$, which implies $\pi(v) = 0$ since ψ is an isomorphism. Hence, $v \in U$, and we have $\ker\phi \subseteq U$. Conversely, if $v \in U$, then $\pi(v) = 0$, so $\phi(v) = \psi(\pi(v)) = 0$. Therefore, $\ker\phi = U$. □

Solution 3E.18

(a) Suppose W is a finite-dimensional subspace of V and $V = U + W$. Let $\pi : V \to V/U$ be the quotient map, and consider the restriction $\pi|_W : W \to V/U$. To see that $\operatorname{im}(\pi|_W) = V/U$, note that for any $v + U \in V/U$, we can write $v = u + w$ for some $u \in U$ and $w \in W$. Then $\pi(v) = \pi(u + w) = \pi(u) + \pi(w) =$

$0 + \pi(w) = \pi(w)$, so $v + U = \pi(w) \in \operatorname{im}(\pi|_W)$. Hence, $\operatorname{im}(\pi|_W) = V/U$. From the fundamental theorem of linear maps, we have

$$\dim W = \dim \ker \pi|_W + \dim \operatorname{im}(\pi|_W) = \dim \ker \pi|_W + \dim V/U \geq \dim V/U.$$

(b) Let $v_1 + U, \ldots, v_m + U$ be a basis of V/U. We claim that $v_1, \ldots, v_m$ are linearly independent. Suppose $\lambda_1 v_1 + \cdots + \lambda_m v_m = 0$ for some scalars $\lambda_1, \ldots, \lambda_m$. Then

$$\lambda_1 v_1 + \cdots + \lambda_m v_m \in U.$$

But $\lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U) = 0$ in V/U, and since $v_1 + U, \ldots, v_m + U$ are linearly independent, it follows that $\lambda_1 = \cdots = \lambda_m = 0$. Hence, $v_1, \ldots, v_m$ are linearly independent.

Let $W = \operatorname{span}\{v_1, \ldots, v_m\}$. Then $\dim W = m = \dim V/U$. For any $v \in V$, we can write

$$v + U = \lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U).$$

This implies that

$$v - (\lambda_1 v_1 + \cdots + \lambda_m v_m) \in U,$$

so $v \in U + W$. Hence, $V = U + W$. To see that $U \cap W = \{0\}$, let $w \in U \cap W$. Then $w \in W$, so $w = \lambda_1 v_1 + \cdots + \lambda_m v_m$ for some scalars $\lambda_1, \ldots, \lambda_m$. But $w \in U$, so $w + U = 0$ in V/U. Hence,

$$\lambda_1(v_1 + U) + \cdots + \lambda_m(v_m + U) = 0$$

in V/U. Since $v_1 + U, \ldots, v_m + U$ are linearly independent, it follows that $\lambda_1 = \cdots = \lambda_m = 0$. Hence, $w = 0$, and we have $U \cap W = \{0\}$. Therefore, $V = U \oplus W$. □

Solution 3E.19 Suppose $S \in \operatorname{Hom}(V/U, W)$ is such that $T = S \circ \pi$. Then for any $u \in U$, we have $T(u) = S(\pi(u)) = S(0) = 0$, so $U \subseteq \ker T$.
Conversely, suppose $U \subseteq \ker T$. Define $S : V/U \to W$ by $S(v + U) = T(v)$. This is well-defined because if $v + U = v' + U$, then $v - v' \in U \subseteq \ker T$, so $T(v) = T(v')$. It is straightforward to check that S is linear and that $T = S \circ \pi$. □

§3F

Solution 3F.1 Let $\phi \in V'$. If $\phi = 0$, then it is the zero map. Otherwise, there exists $v \in V$ such that $\phi(v) \neq 0$, say $\phi(v) = c \neq 0$. Then for any $x \in \mathbb{F}$, we have

$$\phi\left(\frac{x}{c} v\right) = \frac{x}{c}\phi(v) = \frac{x}{c} c = x.$$

Hence, ϕ is surjective. □

Solution 3F.2 Consider the linear functionals on $\mathbb{R}^{[0,1]}$ defined by

$$\phi_1(f) = f(0), \quad \phi_2(f) = f(1), \quad \phi_3(f) = f\left(\frac{1}{2}\right). \qquad \square$$

Solution 3F.3 Since V is finite-dimensional, we can extend v to a basis of V, say $v, v_2, \ldots, v_n$. Define $\phi \in V'$ by setting $\phi(v) = 1$ and $\phi(v_i) = 0$ for $i = 2, \ldots, n$. This defines a linear functional because it is defined on a basis and extended linearly. Clearly, $\phi(v) = 1$. $\square$

Solution 3F.4 Since $U \neq V$, there exists $v_1 \in V$ such that $v_1 \notin U$. Extend a basis of U to a basis of V, say $u_1, \ldots, u_k, v_1, \ldots, v_m$. Define $\phi \in V'$ by setting $\phi(u_i) = 0$ for all $i = 1, \ldots, k$ and $\phi(v_j) = 0$ for all $j = 2, \ldots, m$ but $\phi(v_1) = 1$. This defines a linear functional because it is defined on a basis and extended linearly. Clearly, $\phi(u) = 0$ for all $u \in U$ and $\phi \neq 0$. $\square$

Solution 3F.5 Let $v \in V$. Since $w_1, \ldots, w_m$ is a basis of $\operatorname{im} T$, there exist unique scalars $\phi_1(v), \ldots, \phi_m(v)$ such that

$$Tv = \phi_1(v)w_1 + \cdots + \phi_m(v)w_m.$$

To show that each ϕ_i is a linear functional, let $v, u \in V$ and $a, b \in \mathbb{F}$. Then

$$T(av+bu) = aTv+bTu = a(\phi_1(v)w_1+\cdots+\phi_m(v)w_m)+b(\phi_1(u)w_1+\cdots+\phi_m(u)w_m).$$

By the uniqueness of the representation in terms of the basis $w_1, \ldots, w_m$, we must have

$$\phi_i(av + bu) = a\phi_i(v) + b\phi_i(u) \quad \text{for each } i = 1, \ldots, m.$$

Hence, each ϕ_i is a linear functional on V. $\square$

Solution 3F.6 Suppose $\ker \phi \subseteq \ker \beta$. If $\phi = 0$, then $\ker \phi = V$, and hence $\ker \beta = V$, which implies $\beta = 0$. In this case, we can take $c = 0$ and have $\beta = c\phi$. Now assume $\phi \neq 0$. Pick $v_0 \in V$ such that $\phi(v_0) \neq 0$. Define

$$c = \frac{\beta(v_0)}{\phi(v_0)}.$$

For any $v \in V$, consider

$$v - \frac{\phi(v)}{\phi(v_0)}v_0.$$

This vector is in $\ker \phi$, so it is also in $\ker \beta$. Hence,

$$\beta\left(v - \frac{\phi(v)}{\phi(v_0)}v_0\right) = 0 \implies \beta(v) = \frac{\phi(v)}{\phi(v_0)}\beta(v_0) = c\phi(v).$$

Therefore, $\beta = c\phi$.
Conversely, if $\beta = c\phi$, then clearly $\ker \phi \subseteq \ker \beta$ because if $\phi(v) = 0$, then $\beta(v) = c\phi(v) = 0$. $\square$

Solution 3F.7 Define a map $\Phi : (V_1 \times \cdots \times V_m)' \to V_1' \times \cdots \times V_m'$ by

$$\Phi(\phi) = (\phi_1, \ldots, \phi_m),$$

where $\phi_i \in V_i'$ is defined by $\phi_i(v_i) = \phi(0, \ldots, 0, v_i, 0, \ldots, 0)$, with v_i in the ith position. First, we show that Φ is linear. Let $\phi, \psi \in (V_1 \times \cdots \times V_m)'$ and $a, b \in \mathbb{F}$. Then

$$\Phi(a\phi + b\psi) = ((a\phi + b\psi)_1, \ldots, (a\phi + b\psi)_m) = (a\phi_1 + b\psi_1, \ldots, a\phi_m + b\psi_m) = a\Phi(\phi) + b\Phi(\psi).$$

Next, we show that Φ is bijective. For injectivity, suppose $\Phi(\phi) = 0$. Then $\phi_i = 0$ for all i, which implies $\phi(0, \ldots, 0, v_i, 0, \ldots, 0) = 0$ for all $v_i \in V_i$. By linearity, ϕ must be zero on all of $V_1 \times \cdots \times V_m$, so $\phi = 0$. For surjectivity, given $(\phi_1, \ldots, \phi_m) \in V_1' \times \cdots \times V_m'$, define $\phi \in (V_1 \times \cdots \times V_m)'$ by

$$\phi(v_1, \ldots, v_m) = \phi_1(v_1) + \cdots + \phi_m(v_m),$$

which is linear by linearity of each ϕ_i. Then $\Phi(\phi) = (\phi_1, \ldots, \phi_m)$, so Φ is surjective.

Therefore, Φ is a linear isomorphism, and $(V_1 \times \cdots \times V_m)'$ and $V_1' \times \cdots \times V_m'$ are isomorphic vector spaces. □

Solution 3F.8 To show that Γ and Λ are inverses of each other, we need to check that $\Gamma(\Lambda(a_1, \ldots, a_n)) = (a_1, \ldots, a_n)$ for all $(a_1, \ldots, a_n) \in \mathbb{F}^n$ and $\Lambda(\Gamma(v)) = v$ for all $v \in V$.
First, let $(a_1, \ldots, a_n) \in \mathbb{F}^n$. Then

$$\Gamma(\Lambda(a_1, \ldots, a_n)) = \Gamma(a_1 v_1 + \cdots + a_n v_n) = (\phi_1(a_1 v_1 + \cdots + a_n v_n), \ldots, \phi_n(a_1 v_1 + \cdots + a_n v_n)).$$

By the definition of the dual basis, we have $\phi_i(v_j) = \delta_{ij}$ for all i, j. Therefore,

$$\phi_i(a_1 v_1 + \cdots + a_n v_n) = a_i.$$

Therefore,

$$\Gamma(\Lambda(a_1, \ldots, a_n)) = (a_1, \ldots, a_n).$$

Next, let $v \in V$. Write $v = b_1 v_1 + \cdots + b_n v_n$ for some scalars $b_1, \ldots, b_n$. Then

$$\Lambda(\Gamma(v)) = \Lambda(\phi_1(v), \ldots, \phi_n(v)) = \Lambda(b_1, \ldots, b_n) = b_1 v_1 + \cdots + b_n v_n = v.$$

Hence, Γ and Λ are inverses of each other. □

Solution 3F.9

Axler Note. *Here $p^{(k)}$ denotes the kth derivative of p, with the understanding that the 0th derivative of p is p.*

Let $p \in \mathbb{R}[X]_{\leq m}$. We can write $p(x) = a_0 + a_1 x + \cdots + a_m x^m$ for some scalars $a_0, \ldots, a_m$. Then for each $k = 0, 1, \ldots, m$, we have

$$\begin{aligned}
p(x) &= a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m \\
p'(x) &= a_1 + 2a_2 x + \cdots + m a_m x^{m-1} \\
p''(x) &= 2a_2 + \cdots + m(m-1) a_m x^{m-2} \\
&\vdots \\
p^{(k)}(x) &= k! a_k + \cdots + \frac{m!}{(m-k)!} a_m x^{m-k}
\end{aligned}$$

Then

$$\phi_k(p) = \frac{p^{(k)}(0)}{k!} = a_k.$$

Therefore, $\phi_k(x^j) = \delta_{kj}$ for all $0 \le k, j \le m$, which shows that $\phi_0, \phi_1, \ldots, \phi_m$ is indeed the dual basis of $1, x, \ldots, x^m$. □

Solution 3F.10

(a) Note that $1, x-5, \ldots, (x-5)^m$ is a list of $m+1$ polynomials in $\mathbb{R}[X]_{\le m}$ with differing degrees, and since $\mathbb{R}[X]_{\le m}$ has dimension $m+1$, this list forms a basis of $\mathbb{R}[X]_{\le m}$.

(b) Substitute $y = x - 5$. Then the basis $1, x-5, \ldots, (x-5)^m$ becomes $1, y, \ldots, y^m$. By Exercise 3F9, the dual basis of $1, y, \ldots, y^m$ is $\phi_0, \phi_1, \ldots, \phi_m$, where

$$\phi_k(p) = \frac{p^{(k)}(0)}{k!}.$$

Since $y = x - 5$, then

$$\phi_k(p) = \frac{p^{(k)}(5)}{k!}.$$ □

Solution 3F.11 Let $v \in V$. Since $\phi_1, \ldots, \phi_n$ is the dual basis of $v_1, \ldots, v_n$, we have $\phi_i(v_j) = \delta_{ij}$ for all $1 \le i, j \le n$. Let $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for some scalars $\alpha_1, \ldots, \alpha_n$. Then

$$\begin{aligned}(\psi(v_1)\phi_1 + \cdots + \psi(v_n)\phi_n)(v) &= \psi(v_1)\phi_1(v) + \cdots + \psi(v_n)\phi_n(v)\\ &= \psi(v_1)\alpha_1 + \cdots + \psi(v_n)\alpha_n\\ &= \psi(\alpha_1 v_1 + \cdots + \alpha_n v_n)\\ &= \psi(v)\end{aligned}$$

Since this holds for all $v \in V$, we conclude that

$$\psi = \psi(v_1)\phi_1 + \cdots + \psi(v_n)\phi_n.$$ □

Solution 3F.12

(a) Let $\psi \in W'$. Then for any $v \in V$, we have

$$\begin{aligned}((S+T)'(\psi))(v) &= \psi((S+T)(v))\\ &= \psi(S(v) + T(v))\\ &= \psi(S(v)) + \psi(T(v))\\ &= (S'(\psi))(v) + (T'(\psi))(v)\\ &= (S'(\psi) + T'(\psi))(v).\end{aligned}$$

Since this holds for all $v \in V$, we conclude that

$$(S+T)'(\psi) = S'(\psi) + T'(\psi).$$

Since this holds for all $\psi \in W'$, we have

$$(S+T)' = S' + T'.$$

(b) Let $\psi \in W'$. Then for any $v \in V$, we have

$$\begin{aligned}((\lambda T)'(\psi))(v) &= \psi((\lambda T)(v)) \\ &= \psi(\lambda T(v)) \\ &= \lambda\psi(T(v)) \\ &= (\lambda T'(\psi))(v).\end{aligned}$$

Since this holds for all $v \in V$, we conclude that

$$(\lambda T)'(\psi) = \lambda T'(\psi).$$

Since this holds for all $\psi \in W'$, we have

$$(\lambda T)' = \lambda T'. \qquad \square$$

Solution 3F.13 Let $\psi \in V'$. Let I' denote the dual map of the identity operator I on V. Then for any $v \in V$, we have

$$\begin{aligned}(I'(\psi))(v) &= \psi(I(v)) \\ &= \psi(v).\end{aligned}$$

Since this holds for all $v \in V$, we conclude that

$$I'(\psi) = \psi.$$

Since this holds for all $\psi \in V'$, then I' must be the identity operator on V'. $\square$

Solution 3F.14

(a) Let $(x, y, z) \in \mathbb{R}^3$. Then we have

$$\begin{aligned}(T'(\phi_1))((x,y,z)) &= \phi_1(T(x,y,z)) \\ &= \phi_1(4x+5y+6z, 7x+8y+9z) \\ &= 4x+5y+6z, \\ (T'(\phi_2))((x,y,z)) &= \phi_2(T(x,y,z)) \\ &= \phi_2(4x+5y+6z, 7x+8y+9z) \\ &= 7x+8y+9z.\end{aligned}$$

(b) We can write $T'(\phi_1)$ and $T'(\phi_2)$ as linear combinations of ψ_1, ψ_2, ψ_3 as follows:

$$\begin{aligned}T'(\phi_1) &= 4\psi_1 + 5\psi_2 + 6\psi_3, \\ T'(\phi_2) &= 7\psi_1 + 8\psi_2 + 9\psi_3.\end{aligned}$$

$\square$

Solution 3F.15

(a) Let $p \in \mathbb{R}[X]$. Then we have

$$\begin{aligned}(T'(\phi))(p) &= \phi(T(p)) \\ &= (Tp)'(4) \\ &= (2xp(x) + x^2 p'(x) + p'''(x))\big|_{x=4}. \\ &= 8p(4) + 16p'(4) + p'''(4).\end{aligned}$$

(b) Let $p(x) = x^3$. Then we have

$$\begin{aligned}(T'(\phi))(x^3) &= \phi(T(x^3)) \\ &= \int_0^1 T(x^3)\,dx \\ &= \int_0^1 (x^2 \cdot x^3 + (x^3)'')\,dx \\ &= \int_0^1 (x^5 + 6x)\,dx \\ &= \frac{1}{6} + \frac{18}{6} \\ &= \frac{19}{6}.\end{aligned}$$

□

Solution 3F.16 Suppose $T' = 0$. Then for every $\phi \in W'$ and every $v \in V$, we have

$$(T'(\phi))(v) = \phi(T(v)) = 0.$$

If $T(v) \neq 0$ for some $v \in V$, then $T(v) \in W$. Since W is finite-dimensional, then by Exercise 3F3, there exists $\psi \in W'$ such that $\psi(T(v)) \neq 0$, which contradicts the fact that $(T'(\psi))(v) = \psi(T(v)) = 0$. Hence, $T(v) = 0$ for every $v \in V$, and so $T = 0$.
Conversely, suppose $T = 0$. Then for every $\phi \in W'$ and every $v \in V$, we have

$$(T'(\phi))(v) = \phi(T(v)) = \phi(0) = 0.$$

Hence, $T' = 0$. □

Solution 3F.17 Suppose T is invertible. Then there exists $S \in \mathrm{Hom}(W, V)$ such that $ST = I_V$ and $TS = I_W$. Taking the dual of both sides, we get $T'S' = I_{W'}$ and $S'T' = I_{V'}$, which shows that T' is invertible with inverse S'.
Conversely, suppose T' is invertible. Then T' is an isomorphism. Then

$$\dim V = \dim V' = \dim W' = \dim W.$$

By injectivity of T', we have $\ker T' = \{0\}$. Since $\ker T' = (\mathrm{im}\, T)^0$, we have $(\mathrm{im}\, T)^0 = \{0\}$ so that $\mathrm{im}\, T = W$. Hence, T is surjective. Since $\dim V = \dim W$, we conclude that T is injective as well. Therefore, T is invertible. □

Solution 3F.18 Let $\Phi : \mathrm{Hom}(V, W) \to \mathrm{Hom}(W', V')$ be the map defined by $\Phi(T) = T'$. We need to show that Φ is linear, injective, and surjective.
Linearity: For $T_1, T_2 \in \mathrm{Hom}(V, W)$ and $a \in \mathbb{F}$, we have

$$\Phi(T_1 + T_2) = (T_1 + T_2)' = T_1' + T_2' = \Phi(T_1) + \Phi(T_2),$$

and

$$\Phi(aT_1) = (aT_1)' = aT_1' = a\Phi(T_1).$$

Hence, Φ is linear.

To see that Φ is injective, suppose $\Phi(T) = T' = 0$. By Exercise 3F16, this implies $T = 0$. Hence, Φ is injective.

To see that Φ is surjective, note that since V and W are finite-dimensional, $\dim \mathrm{Hom}(V, W) = \dim V \cdot \dim W = \dim W' \cdot \dim V' = \dim \mathrm{Hom}(W', V')$. A linear map between finite-dimensional vector spaces of the same dimension that is injective is also surjective. Hence, Φ is surjective and therefore an isomorphism. □

Solution 3F.19 By definition, the annihilator U^0 of U is the set of all linear functionals $\phi \in V'$ such that $\phi(u) = 0$ for all $u \in U$. Equivalently, this means that $U \subseteq \ker \phi$. □

Solution 3F.20

Manual Remark. *This says that* $U = (U^0)^0$.

Let $S = \{v \in V : \phi(v) = 0 \text{ for every } \phi \in U^0\}$. We need to show that $S = U$.

First, if $v \in U$, then for any $\phi \in U^0$, we have $\phi(v) = 0$ by definition of the annihilator. Hence, $v \in S$, so $U \subseteq S$.

Conversely, suppose $v \in S$. If $v \notin U$, then by the construction done in Exercise 3F4 with $v_1 = v$, there exists a linear functional $\phi \in V'$ such that $\phi(u) = 0$ for all $u \in U$ and $\phi(v) \neq 0$. But this ϕ is in U^0, which contradicts $v \in S$. Hence, $v \in U$, and therefore $S \subseteq U$.

Combining both inclusions, we get $S = U$, as desired. □

Solution 3F.21

(a) Suppose $W^0 \subseteq U^0$. If $u \in U$, then for any $\phi \in W^0$, we have $\phi(u) = 0$ because $\phi \in U^0$. Hence, by Exercise 3F20, we have $u \in W$, so $U \subseteq W$. Conversely, if $U \subseteq W$ and $\phi \in W^0$, then $\phi(w) = 0$ for all $w \in W$, and in particular for all $u \in U$. Hence, $\phi \in U^0$, so $W^0 \subseteq U^0$.

(b) Suppose $W^0 = U^0$. Then $W^0 \subseteq U^0$ and $U^0 \subseteq W^0$. By part (a), this implies that $U \subseteq W$ and $W \subseteq U$. Hence, $U = W$. Conversely, if $U = W$, then clearly $U^0 = W^0$.

□

Solution 3F.22

(a) Let $\phi \in (U + W)^0$. Then for any $u \in U$ and $w \in W$, we have $\phi(u + w) = 0$. In particular, taking $w = 0$ gives $\phi(u) = 0$ for all $u \in U$, so $\phi \in U^0$. Similarly, taking $u = 0$ gives $\phi(w) = 0$ for all $w \in W$, so $\phi \in W^0$. Hence, $\phi \in U^0 \cap W^0$, and we have shown that $(U + W)^0 \subseteq U^0 \cap W^0$.

Conversely, let $\phi \in U^0 \cap W^0$. Then for any $u \in U$ and $w \in W$, we have $\phi(u) = 0$ and $\phi(w) = 0$, so $\phi(u + w) = \phi(u) + \phi(w) = 0$. Hence, $\phi \in (U + W)^0$, and we have shown that $U^0 \cap W^0 \subseteq (U + W)^0$. Combining both inclusions, we get $(U + W)^0 = U^0 \cap W^0$.

(b) From part (a), we know that $(U + W)^0 = U^0 \cap W^0$. Then

$$\dim(U + W)^0 = \dim V - \dim(U + W).$$

Applying the dimension formula for a sum of subspaces, we have

$$\dim(U^0 + W^0) = \dim U^0 + \dim W^0 - \dim(U^0 \cap W^0).$$

Since $(U + W)^0 = U^0 \cap W^0$, we have $\dim(U^0 \cap W^0) = \dim(U + W)^0 = \dim V - \dim(U + W)$. Therefore,

$$\dim(U^0 + W^0) = \dim U^0 + \dim W^0 - (\dim V - \dim(U + W)).$$

Using the fact that $\dim U^0 = \dim V - \dim U$ and $\dim W^0 = \dim V - \dim W$, we get

$$\begin{aligned}
\dim(U^0 + W^0) &= \dim U^0 + \dim W^0 - \dim(U^0 \cap W^0) \\
&= (\dim V - \dim U) + (\dim V - \dim W) - (\dim V - \dim(U + W)) \\
&= \dim V - \dim U + \dim V - \dim W - \dim V + \dim(U + W) \\
&= \dim V - (\dim U + \dim W - \dim(U + W)) \\
&= \dim V - \dim(U \cap W).
\end{aligned}$$

Hence, $\dim(U^0 + W^0) = \dim(U \cap W)^0$. To see equality, we just need to prove one inclusion. If $\phi = \phi_U + \phi_W \in U^0 + W^0$, then for any $v \in U \cap W$, we have $\phi(v) = \phi_U(v) + \phi_W(v) = 0 + 0 = 0$, so $\phi \in (U \cap W)^0$. Therefore, $U^0 + W^0 \subseteq (U \cap W)^0$. Since the dimensions are equal, we conclude that $(U \cap W)^0 = U^0 + W^0$.

□

Solution 3F.23 For ease of notation, let A, B, C be the sets in the exercise, i.e.,

$$\begin{aligned}
A &= \operatorname{span}(\phi_1, \ldots, \phi_m), \\
B &= ((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0, \\
C &= \{\phi \in V' : (\ker \phi_1) \cap \cdots \cap (\ker \phi_m) \subseteq \ker \phi\}.
\end{aligned}$$

We will show that $A = B$ and $B = C$. Starting off with $B = C$, note that by definition,

$$\begin{aligned}
B &= ((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0 \\
&= \{\phi \in V' : (\ker \phi_1) \cap \cdots \cap (\ker \phi_m) \subseteq \ker \phi\} = C,
\end{aligned}$$

where the reformulation of the annihilator subspace was shown in Exercise 3F19.
To show that $A = B$, we may use Exercise 3F22 (b) repeatedly to rewrite B as

$$B = (\ker \phi_1)^0 + \cdots + (\ker \phi_m)^0.$$

By Exercise 3F6, we have $(\ker \phi_i)^0 = \operatorname{span}(\phi_i)$ for each $i = 1, \ldots, m$. Therefore,

$$B = (\ker \phi_1)^0 + \cdots + (\ker \phi_m)^0 = \operatorname{span}(\phi_1) + \cdots + \operatorname{span}(\phi_m) = \operatorname{span}(\phi_1, \ldots, \phi_m) = A.$$

We conclude that $A = B$. □

Solution 3F.24

(a) Suppose $v_1, \ldots, v_m$ spans V. If $\Gamma(\phi) = 0$, then $\phi(v_i) = 0$ for all $i = 1, \ldots, m$. Since $v_1, \ldots, v_m$ spans V, it follows that $\phi(v) = 0$ for all $v \in V$, i.e., $\phi = 0$. Hence, Γ is injective.

Conversely, suppose Γ is injective. If $v_1, \ldots, v_m$ does not span V, then by Exercise 3F4, there exists a nonzero $\phi \in V'$ that vanishes on $v_1, \ldots, v_m$. But then $\Gamma(\phi) = 0$, contradicting the injectivity of Γ. Hence, $v_1, \ldots, v_m$ must span V.

(b) Suppose $v_1, \ldots, v_m$ is linearly independent. Extend this list to a basis $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ of V. Define linear functionals $\phi_1, \ldots, \phi_m \in V'$ by setting $\phi_i(v_j) = \delta_{ij}$ for $i, j = 1, \ldots, m$ and $\phi_i(v_j) = 0$ for $j = m+1, \ldots, n$. Then for any $(a_1, \ldots, a_m) \in \mathbb{F}^m$, the linear functional $\phi = a_1\phi_1 + \cdots + a_m\phi_m$ satisfies $\Gamma(\phi) = (a_1, \ldots, a_m)$. Hence, Γ is surjective.

Conversely, suppose Γ is surjective. If $v_1, \ldots, v_m$ is not linearly independent, then there exist scalars $a_1, \ldots, a_m$, not all zero, such that $a_1v_1 + \cdots + a_mv_m = 0$. Consider some index j such that $a_j \neq 0$. Since Γ is surjective, there exists $\phi \in V'$ such that $\phi(v_j) = 1$ and $\phi(v_i) = 0$ for all $i \neq j$. Then

$$0 = \phi(a_1v_1 + \cdots + a_mv_m) = a_1\phi(v_1) + \cdots + a_m\phi(v_m) = a_j \neq 0,$$

a contradiction. Hence, $v_1, \ldots, v_m$ must be linearly independent.

□

Solution 3F.25

(a) Suppose $\phi_1, \ldots, \phi_m$ spans V'. If $\Gamma(v) = 0$, then $\phi_i(v) = 0$ for all $i = 1, \ldots, m$. Since $\phi_1, \ldots, \phi_m$ spans V', then by Exercise 3F3, it follows that $\psi(v) = 0$ for all $\psi \in V'$, i.e., $v = 0$. Hence, Γ is injective.

Conversely, suppose Γ is injective. Then $\ker \Gamma = \{0\}$. By definition, we have

$$\ker \Gamma = \ker \phi_1 \cap \cdots \cap \ker \phi_m.$$

By Exercise 3F23, we have

$$\operatorname{span}(\phi_1, \ldots, \phi_m) = ((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0 = (\ker \Gamma)^0 = \{\psi \in V' : \ker \Gamma \subseteq \ker \psi\} = V'.$$

Hence, $\phi_1, \ldots, \phi_m$ spans V'.

(b) Suppose $\phi_1, \ldots, \phi_m$ is linearly independent. Then $\dim \operatorname{span}(\phi_1, \ldots, \phi_m) = m$. By the fundamental theorem of linear maps and Exercise 3F23, we have

$$\dim \operatorname{im} \Gamma = \dim V - \dim \ker \Gamma = \dim V - \dim((\ker \phi_1) \cap \cdots \cap (\ker \phi_m)) = m.$$

Since $\operatorname{im} \Gamma$ is a subspace of $\mathbb{F}^m$ and has dimension m, we conclude that $\operatorname{im} \Gamma = \mathbb{F}^m$, so Γ is surjective.

Conversely, suppose Γ is surjective. Then $\operatorname{im} \Gamma = \mathbb{F}^m$, so $\dim \operatorname{im} \Gamma = m$. By the fundamental theorem of linear maps, we have

$$\dim \operatorname{im} \Gamma = \dim V - \dim \ker \Gamma = \dim V - \dim((\ker \phi_1) \cap \cdots \cap (\ker \phi_m)).$$

By Exercise 3F23, we have

$$\operatorname{span}(\phi_1, \dots, \phi_m) = ((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0,$$

so $\dim \operatorname{span}(\phi_1, \dots, \phi_m) = \dim V - \dim((\ker \phi_1) \cap \cdots \cap (\ker \phi_m)) = m$. Hence, $\phi_1, \dots, \phi_m$ is linearly independent.

□

Solution 3F.26

Manual Remark. *This says* $\Omega = (\Omega^0)^0$.

Let $S = \{v \in V : \phi(v) = 0 \text{ for every } \phi \in \Omega\}$. We need to show that $\Omega = S^0$. Let $\phi_1, \dots, \phi_m$ be a basis of Ω. Then $S = (\ker \phi_1) \cap \cdots \cap (\ker \phi_m)$. By Exercise 3F23, we have

$$\Omega = \operatorname{span}(\phi_1, \dots, \phi_m) = ((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0 = S^0. \qquad \square$$

Solution 3F.27 Let $W = \{p \in \mathbb{R}[X]_{\le 5} : p(8) = 0\}$. We need to show that $\operatorname{im} T = W$. First, if $p \in \operatorname{im} T$, then there exists $q \in \mathbb{R}[X]_{\le 5}$ such that $T(q) = p$. Since $\phi \in \ker T'$, we have

$$\phi(T(q)) = (T'(\phi))(q) = 0.$$

But $\phi(T(q)) = T(q)(8) = p(8)$, so $p(8) = 0$. Hence, $p \in W$, and we have shown that $\operatorname{im} T \subseteq W$.

Now we compare the dimensions of $\operatorname{im} T$ and W. Recall that $(\operatorname{im} T)^0 = \ker T' = \operatorname{span}(\phi)$, so $\dim(\operatorname{im} T)^0 = 1$. Since $\mathbb{R}[X]_{\le 5}$ has dimension 6, we have $\dim \operatorname{im} T = 6 - 1 = 5$. On the other hand, W is the set of all polynomials in $\mathbb{R}[X]_{\le 5}$ that vanish at 8. Using the fundamental theorem of linear maps and the fact that $\phi \neq 0$ is surjective by Exercise 3F1, we have $\dim W = \dim \mathbb{R}[X]_{\le 5} - 1 = 5$. Hence, $\operatorname{im} T$ and W are subspaces of $\mathbb{R}[X]_{\le 5}$ with the same dimension, and $\operatorname{im} T \subseteq W$. We conclude that $\operatorname{im} T = W$, as desired. □

Solution 3F.28 Since $\phi_1, \dots, \phi_m$ is linearly independent, we have $\dim \operatorname{span}(\phi_1, \dots, \phi_m) = m$. By Exercise 3F23, we have

$$\operatorname{span}(\phi_1, \dots, \phi_m) = ((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0.$$

Hence, $\dim((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0 = m$. Since V is finite-dimensional, we have

$$\dim((\ker \phi_1) \cap \cdots \cap (\ker \phi_m)) = \dim V - \dim((\ker \phi_1) \cap \cdots \cap (\ker \phi_m))^0 = \dim V - m,$$

as desired. □

Solution 3F.29

(a) From $\ker T' = \operatorname{span}(\phi)$ and $\ker T' = (\operatorname{im} T)^0$, we have by Exercise 3F20 that

$$\operatorname{im} T = ((\operatorname{im} T)^0)^0 = (\ker T')^0 = (\operatorname{span}(\phi))^0 = \ker \phi,$$

where the last equality follows as such:

$$v \in \ker \phi \iff \phi(v) = 0 \iff \alpha\phi(v) = 0 \text{ for all } \alpha \in \mathbb{F} \iff v \in (\operatorname{span}(\phi))^0.$$

(b) Note that $\operatorname{im} T' = \operatorname{span}(\psi)$ and $\operatorname{im} T' = (\ker T)^0$. Then $(\ker T)^0 = \operatorname{span}(\psi)$. By Exercise 3F20, we have

$$\ker T = ((\ker T)^0)^0 = (\operatorname{span}(\psi))^0 = \ker \psi,$$

where the last equality follows as such:

$$v \in \ker \psi \iff \psi(v) = 0 \iff \alpha\psi(v) = 0 \text{ for all } \alpha \in \mathbb{F} \iff v \in (\operatorname{span}(\psi))^0.$$

□

Solution 3F.30 Let $\phi_1, \ldots, \phi_n$ be a basis of V'. From Exercise 3F25, there exists a map $\Gamma : V \to \mathbb{F}^n$ that is an isomorphism, since $\phi_1, \ldots, \phi_n$ is a basis for V'. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. Since Γ is an isomorphism, there exist $v_1, \ldots, v_n \in V$ such that $\Gamma(v_i) = e_i$ for each $i = 1, \ldots, n$. We claim that $v_1, \ldots, v_n$ is a basis of V whose dual basis is $\phi_1, \ldots, \phi_n$.
To see that $v_1, \ldots, v_n$ is a basis of V, we show linear independence. Suppose $a_1, \ldots, a_n$ are scalars such that $a_1 v_1 + \cdots + a_n v_n = 0$. Applying Γ to both sides gives

$$a_1 e_1 + \cdots + a_n e_n = \Gamma(a_1 v_1 + \cdots + a_n v_n) = \Gamma(0) = 0.$$

Since $e_1, \ldots, e_n$ is a basis of $\mathbb{F}^n$, it follows that $a_1 = \cdots = a_n = 0$. Hence, $v_1, \ldots, v_n$ is linearly independent. Since $\dim V = n$, we conclude that $v_1, \ldots, v_n$ is a basis of V.
To see that the dual basis of $v_1, \ldots, v_n$ is $\phi_1, \ldots, \phi_n$, we need to show that $\phi_i(v_j) = \delta_{ij}$ for each $i, j = 1, \ldots, n$. Since $\Gamma(v_j) = e_j$, we have

$$\phi_i(v_j) = (\Gamma(v_j))_i = (e_j)_i = \delta_{ij},$$

as desired. □

Solution 3F.31

(a) Note that if $\phi \in \ker i'$, then $i'(\phi) = 0$, so $\phi(i(u)) = 0$ for all $u \in U$. Since $i(u) = u$, we have $\phi(u) = 0$ for all $u \in U$, so $\phi \in U^0$. Hence, $\ker i' \subseteq U^0$. Conversely, if $\phi \in U^0$, then $\phi(u) = 0$ for all $u \in U$, so $i'(\phi) = 0$. Hence, $U^0 \subseteq \ker i'$, and we conclude that $\ker i' = U^0$.

(b) To see $\operatorname{im} i' \subseteq U'$, note that if $\phi \in V'$, then $i'(\phi) \in U'$ since $i'(\phi)(u) = \phi(i(u)) = \phi(u)$ for all $u \in U$. Hence, $\operatorname{im} i' \subseteq U'$.

Since V is finite-dimensional, we have $\dim V' = \dim V$. By the fundamental theorem of linear maps, we have

$$\dim \operatorname{im} i' = \dim V' - \dim \ker i' = \dim V - \dim U^0.$$

By the annihilator dimension formula, we have $\dim U^0 = \dim V - \dim U$. Hence,

$$\dim \operatorname{im} i' = \dim V - (\dim V - \dim U) = \dim U = \dim U'.$$

Since $\operatorname{im} i'$ is a subspace of U' and has the same dimension, we conclude that $\operatorname{im} i' = U'$, as desired.

(c) From part (a), we have $\ker i' = U^0$. Then we know that V'/U^0 is isomorphic to $\operatorname{im} i'$. Then the induced map $\tilde{i'} : V'/U^0 \to \operatorname{im} i'$ defined by $\tilde{i'}(\phi + U^0) = i'(\phi)$ is an isomorphism. Since $\operatorname{im} i' = U'$ by part (b), we conclude that i' induces an isomorphism from V'/U^0 onto U', as desired.

□

Solution 3F.32

Axler Note. *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.*

(a) To see that Λ is linear, let $v, w \in V$ and $a \in \mathbb{F}$. Then for any $\phi \in V'$, we have

$$(\Lambda(v + w))(\phi) = \phi(v + w) = \phi(v) + \phi(w) = (\Lambda v)(\phi) + (\Lambda w)(\phi),$$
$$(\Lambda(av))(\phi) = \phi(av) = a\phi(v) = a(\Lambda v)(\phi).$$

Hence, Λ is linear.

(b) Let $v \in V$ and $\phi \in V'$. Then

$$\begin{aligned}((T'' \circ \Lambda)(v))(\phi) &= (T''(\Lambda v))(\phi)\\ &= (\Lambda v)(T'(\phi))\\ &= T'(\phi)(v)\\ &= \phi(T(v))\\ &= (\Lambda(T(v)))(\phi) = ((\Lambda \circ T)(v))(\phi).\end{aligned}$$

Hence, $T'' \circ \Lambda = \Lambda \circ T$.

(c) Since V is finite-dimensional, we have $\dim V = \dim V'$. Then $\dim V'' = \dim(V')' = \dim V' = \dim V$. By the fundamental theorem of linear maps, we have

$$\dim \operatorname{im} \Lambda = \dim V - \dim \ker \Lambda.$$

To see that Λ is injective, suppose $\Lambda v = 0$ for some $v \in V$. Then for any $\phi \in V'$, we have $\phi(v) = (\Lambda v)(\phi) = 0$. Since this holds for all $\phi \in V'$ by Exercise 3F3, it follows that $v = 0$. Hence, Λ is injective, so $\ker \Lambda = \{0\}$. Therefore, $\dim \operatorname{im} \Lambda = \dim V - 0 = \dim V$. Since $\operatorname{im} \Lambda$ is a subspace of V'' and has the same dimension as V'', we conclude that $\operatorname{im} \Lambda = V''$, so Λ is surjective. Combining injectivity and surjectivity, we conclude that Λ is an isomorphism from V onto V'', as desired.

□

Solution 3F.33

(a) Suppose $\pi'(\phi) = 0$ for some $\phi \in (V/U)'$. Then $\phi(\pi(v)) = 0$ for all $v \in V$. Since $\pi(v) = v + U$, we have $\phi(v + U) = 0$ for all $v \in V$. Since this holds for all $v \in V$ by surjectivity of π, it follows that $\phi = 0$. Hence, π' is injective.

(b) To see $\operatorname{im}\pi' \subseteq U^0$, note that if $\phi \in (V/U)'$, then for any $u \in U$, we have $\pi'(\phi)(u) = \phi(\pi(u)) = \phi(0 + U) = 0$. Hence, $\operatorname{im}\pi' \subseteq U^0$.

To see the other inclusion, let $\psi \in U^0$. We want to find $\phi \in (V/U)'$ such that $\pi'(\phi) = \psi$. Define $\phi : V/U \to \mathbb{F}$ by $\phi(v + U) = \psi(v)$ for each $v \in V$. We need to check that ϕ is well-defined. If $v + U = w + U$, then $v - w \in U$, so $\psi(v - w) = 0$, which implies that $\psi(v) = \psi(w)$. Hence, ϕ is well-defined. It is also linear since for any $a, b \in \mathbb{F}$ and $v, w \in V$, we have

$$\begin{aligned}\phi(a(v + U) + b(w + U)) &= \phi((av + bw) + U) = \psi(av + bw)\\ &= a\psi(v) + b\psi(w) = a\phi(v + U) + b\phi(w + U).\end{aligned}$$

Finally, we have $\pi'(\phi)(v) = \phi(\pi(v)) = \phi(v + U) = \psi(v)$ for all $v \in V$, so $\pi'(\phi) = \psi$. Hence, $U^0 \subseteq \operatorname{im}\pi'$, and we conclude that $\operatorname{im}\pi' = U^0$.

(c) From part (a), we have π' is injective. From part (b), we have $\operatorname{im}\pi' = U^0$. Then the induced map $\tilde{\pi}' : (V/U)' \to U^0$ defined by $\tilde{\pi}'(\phi) = \pi'(\phi)$ is an isomorphism. Hence, π' is an isomorphism from $(V/U)'$ onto U^0, as desired.

□

Chapter 4

§4

Solution 4.1 In this problem, let $z = a + bi$ and $w = c + di$ with $a, b, c, d \in \mathbb{R}$. Then

(a) $z + \overline{z} = (a + bi) + (a - bi) = 2a = 2\operatorname{Re} z$.

(b) $z - \overline{z} = (a + bi) - (a - bi) = 2bi = 2(\operatorname{Im} z)i$.

(c) $z\overline{z} = (a + bi)(a - bi) = a^2 + b^2 = |z|^2$.

(d) $\overline{w + z} = \overline{(c + di) + (a + bi)} = \overline{(a + c) + (b + d)i} = (a + c) - (b + d)i = (a - bi) + (c - di) = \overline{w} + \overline{z}$ and $\overline{wz} = \overline{(c + di)(a + bi)} = \overline{(ac - bd) + (ad + bc)i} = (ac - bd) - (ad + bc)i = (c - di)(a - bi) = \overline{w}\overline{z}$.

(e) $\overline{\overline{z}} = \overline{a - bi} = a + bi = z$.

(f) $|\operatorname{Re} z| = |a| \leq \sqrt{a^2 + b^2} = |z|$ and $|\operatorname{Im} z| = |b| \leq \sqrt{a^2 + b^2} = |z|$.

(g) $|\overline{z}| = |a - bi| = \sqrt{a^2 + b^2} = |z|$.

(h) $|wz| = |(c+di)(a+bi)| = |(ac-bd)+(ad+bc)i| = \sqrt{(ac - bd)^2 + (ad + bc)^2} = \sqrt{(a^2 + b^2)(c^2 + d^2)} = |a + bi||c + di| = |z||w|$. □

Solution 4.2 Let $w, z \in \mathbb{C}$. By the triangle inequality, we have

$$|w| = |(w - z) + z| \leq |w - z| + |z|.$$

Rearranging gives $|w| - |z| \leq |w - z|$. Similarly,

$$|z| = |(z - w) + w| \leq |z - w| + |w| = |w - z| + |w|,$$

which gives $|z| - |w| \leq |w - z|$, or equivalently $-|w - z| \leq |w| - |z|$. Combining these inequalities, we get

$$||w| - |z|| \leq |w - z|. \qquad \square$$

Solution 4.3 Let $v \in V$. Then there exist $s, t \in \mathbb{R}$ such that

$$\phi(v) = s + ti.$$

Then

$$\sigma(v) = \operatorname{Re}\phi(v) = s \quad \text{and} \quad \sigma(iv) = \operatorname{Re}\phi(iv) = \operatorname{Re}(i(s + ti)) = \operatorname{Re}(-t + si) = -t.$$

Therefore,

$$\sigma(v) - i\sigma(iv) = s - i(-t) = s + ti = \phi(v). \qquad \square$$

Solution 4.4 No, the set is not a subspace of $\mathbb{F}[X]$. To see this, consider two polynomials $p, q \in \mathbb{F}[X]$ with $p(x) = x^m$ and $q(x) = -x^m + x^{m-1}$. Then $p + q = x^{m-1}$, which has degree $m - 1$ and is not in the set. Therefore, the set is not closed under addition and hence not a subspace. $\square$

Solution 4.5 No, the set is not a subspace of $\mathbb{F}[X]$. To see this, consider two polynomials $p, q \in \mathbb{F}[X]$ with $p(x) = x^2$ and $q(x) = x - x^2$. Then $p + q = x$, which has degree 1 and is not even. Therefore, the set is not closed under addition and hence not a subspace. $\square$

Solution 4.6 Consider the polynomial

$$p(x) = (x - \lambda_1)^{n-m+1}(x - \lambda_2) \cdots (x - \lambda_m).$$

This polynomial has degree n, vanishes at $\lambda_1, \ldots, \lambda_m$, and has no other zeros. Therefore, it satisfies the required conditions. $\square$

Solution 4.7 Consider the linear map $T : \mathbb{F}[X]_{\leq m} \to \mathbb{F}^{m+1}$ defined by

$$T(p) = (p(z_1), \ldots, p(z_{m+1})).$$

Since $\dim \mathbb{F}[X]_{\leq m} = m + 1$ and $\dim \mathbb{F}^{m+1} = m + 1$, it suffices to show that T is injective to conclude that it is also surjective. Suppose $T(p) = 0$, i.e., $p(z_k) = 0$ for all $k = 1, \ldots, m + 1$. Since p has degree at most m and has $m + 1$ distinct zeros, it must be the zero polynomial. Therefore, T is injective, and hence bijective. This shows that for any $(w_1, \ldots, w_{m+1}) \in \mathbb{F}^{m+1}$, there exists a unique $p \in \mathbb{F}[X]_{\leq m}$ such that $T(p) = (w_1, \ldots, w_{m+1})$, which is exactly the desired polynomial. $\square$

Solution 4.8 Suppose p has m distinct zeros, say $z_1, \ldots, z_m$. If p and p' have a common zero z_k, then z_k would be a multiple zero of p, contradicting the assumption that all zeros are distinct. Hence, p and p' have no zeros in common.
Conversely, suppose p and p' have no zeros in common. If p had a multiple zero z, then $p(z) = 0$ and $p'(z) = 0$, which would contradict the assumption. Since p has degree m, it can have at most m zeros by the fundamental theorem of algebra. Therefore, p must have exactly m distinct zeros. □

Solution 4.9

Appendix. *An alternate proof is given at the end of this solution.*

Let $p(x)$ be a polynomial of odd degree with real coefficients. Since non-real zeros come in pairs, there are an even number of non-real zeros. Therefore, there must be at least one real zero.

This solution comes from a calculus perspective, using the intermediate value theorem.

blanketism. Let $p(x)$ be a polynomial of odd degree with real coefficients. As $x \to \infty$, $p(x) \to \infty$ or $p(x) \to -\infty$ depending on the leading coefficient. Similarly, as $x \to -\infty$, $p(x) \to -\infty$ or $p(x) \to \infty$. By the intermediate value theorem, since $p(x)$ is continuous and takes on both positive and negative values, there must exist some $x \in \mathbb{R}$ such that $p(x) = 0$. Therefore, p has a real zero. □

□

Solution 4.10 Let $p \in \mathbb{R}[X]$. Since $p(x) - p(3)$ has a zero at $x = 3$, we can factor it as $(x-3)q(x)$ for some polynomial $q \in \mathbb{R}[X]$. Then for $x \neq 3$,

$$(Tp)(x) = \frac{p(x) - p(3)}{x - 3} = q(x),$$

and differentiating p gives $p'(3) = q(3)$. Since $Tp(3) = p'(3) = q(3)$, we have $Tp(x) = q(x)$ for all $x \in \mathbb{R}$. Therefore, $Tp \in \mathbb{R}[X]$.
Let $p_1, p_2 \in \mathbb{R}[X]$ and $\alpha, \beta \in \mathbb{R}$. Let $p_1(x) - p_1(3) = (x-3)q_1(x)$ and $p_2(x) - p_2(3) = (x-3)q_2(x)$ for some $q_1, q_2 \in \mathbb{R}[X]$. Then

$$(\alpha p_1 + \beta p_2)(x) = (\alpha p_1(3) + \beta p_2(3)) + (x-3)(\alpha q_1(x) + \beta q_2(x)).$$

Hence, $T(\alpha p_1 + \beta p_2)(x) = \alpha q_1(x) + \beta q_2(x) = \alpha T p_1(x) + \beta T p_2(x)$ for all $x \in \mathbb{R}$. Therefore, T is a linear map. □

Solution 4.11

Appendix. *An alternate proof is given at the end of this solution.*

Let $p \in \mathbb{C}[X]$ have the form $p(z) = a_0 + a_1 z + \cdots + a_n z^n$ with $a_k \in \mathbb{C}$. Then

$$\overline{p(\overline{z})} = \overline{\sum_{i=0}^{n} a_i \overline{z}^i} = \sum_{i=0}^{n} \overline{a}_i z^i.$$

Therefore,

$$q(z) = p(z)\overline{p(\overline{z})} = \left(\sum_{i=0}^{n} a_i z^i\right)\left(\sum_{j=0}^{n} \overline{a}_j z^j\right) = \sum_{k=0}^{2n}\left(\sum_{i+j=k} a_i \overline{a}_j\right) z^k.$$

Let b_m denote the coefficient of z^m in $q(z)$. Then

$$b_m = \sum_{i+j=m} a_i \overline{a}_j = \sum_{i+j=m} \overline{a}_j a_i = \overline{b_m}.$$

Hence, $b_m \in \mathbb{R}$ for each $m = 0, 1, \ldots, 2n$. Therefore, q is a polynomial with real coefficients.

This solution examines the roots of the polynomial p and how they contribute to the polynomial q.

blanketism. If $p = 0$, then $q = 0$ is a polynomial with real coefficients. Otherwise, let $n = \deg p$. Then we may write

$$p(z) = \alpha(z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_n),$$

where $\alpha \in \mathbb{C}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ are the roots of p. Then

$$q(z) = p(z)\overline{p(\overline{z})} = \alpha(z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_n)\cdot\overline{\alpha}(z - \overline{\lambda_1})(z - \overline{\lambda_2})\cdots(z - \overline{\lambda_n}).$$

Note that the product contains the term $\alpha\overline{\alpha} = |\alpha|^2 \in \mathbb{R}$, as well as the pairs

$$(z - \lambda_i)(z - \overline{\lambda_i}) = z^2 - 2\,\mathrm{Re}(\lambda_i)z + |\lambda_i|^2.$$

Each such pair contributes a quadratic polynomial with real coefficients. Since the product of polynomials with real coefficients also has real coefficients, it follows that $q(z)$ is a polynomial with real coefficients. □

□

Solution 4.12 Let $p(x_k) = y_k \in \mathbb{R}$ for each $k = 0, 1, \ldots, m$. By Exercise 4.7 with $\mathbb{F} = \mathbb{R}$, there exists a polynomial $q \in \mathbb{R}[X]_{\leq m}$ that satisfies $q(x_k) = y_k$ for each $k = 0, 1, \ldots, m$. Note that this polynomial must be the same as the original $p \in \mathbb{C}[X]_{\leq m}$ because a polynomial of degree at most m is uniquely determined by its values at $m+1$ distinct points. Therefore, all coefficients of p are real. □

Solution 4.13

(a) We first show that U is a subspace of $\mathbb{F}[X]$. Clearly, $0 \in U$ since $0 = p \cdot 0$. If $u_1, u_2 \in U$, then $u_1 = pq_1$ and $u_2 = pq_2$ for some $q_1, q_2 \in \mathbb{F}[X]$. For any scalars $\alpha, \beta \in \mathbb{F}$, we have $\alpha u_1 + \beta u_2 = \alpha p q_1 + \beta p q_2 = p(\alpha q_1 + \beta q_2) \in U$. Therefore, U is closed under addition and scalar multiplication, and hence U is a subspace of $\mathbb{F}[X]$.

Now let $r \in \mathbb{F}[X]$. By the division algorithm for polynomials, there exist unique polynomials $q_0 \in \mathbb{F}[X]$ and $s \in \mathbb{F}[X]$ with $\deg s < \deg p$ such that $r = pq_0 + s$. This shows that every polynomial in $\mathbb{F}[X]$ can be written as an element of U plus

a polynomial of degree less than $\deg p$. Consider the map $T : \mathbb{F}[X]_{\leq \deg p-1} \to \mathbb{F}[X]/U$ defined by $T(s) = s + U$. This map is linear and surjective, and $\ker T$ is trivial since if $s + U = U$, then $s \in U$, which implies $s = pq$ for some $q \in \mathbb{F}[X]$. Then

$$\deg s = \deg p + \deg q < \deg p.$$

Therefore, $q = 0$, hence $s = 0$, and T is an isomorphism. We conclude that $\dim \mathbb{F}[X]/U = \dim \mathbb{F}[X]_{\leq \deg p-1} = \deg p$.

(b) A basis of $\mathbb{F}[X]/U$ is given by the images of the standard basis of $\mathbb{F}[X]_{\leq \deg p-1}$ under T, namely $\{1 + U, x + U, \ldots, x^{\deg p-1} + U\}$.

□

Solution 4.14

Manual Remark. *In part (a), one may also observe that a multiple of p and q must have* $\deg p + \deg q$*, but* $\deg rp = \deg r + \deg p \leq n - 1 + m$.

(a) To show that T is injective, suppose that $T(r, s) = rp + sq = 0$ for some $(r, s) \in \mathbb{C}[X]_{\leq n-1} \times \mathbb{C}[X]_{\leq m-1}$. Then $rp = -sq$ implies rp is a polynomial multiple of q, so every zero of q must be a zero of rp. Because p and q have no zeros in common, every zero of q must be a zero of r. But $\deg r < \deg q$, so it must be that $r = 0$. Because $r = 0$, we have $-sq = 0$. Then since q is nonconstant, it must be that $s = 0$. Therefore, if $(r, s) \in \ker T$, we have $(r, s) = 0$, so T is injective.

(b) To show that T is surjective, note that $\dim(\mathbb{C}[X]_{\leq n-1} \times \mathbb{C}[X]_{\leq m-1}) = n + m = \dim \mathbb{C}[X]_{\leq m+n-1}$. Since T is a linear map between vector spaces of the same finite dimension, and we have already shown that T is injective, it follows that T is also surjective.

(c) Since T is surjective, for the polynomial $1 \in \mathbb{C}[X]_{\leq m+n-1}$, there exist $r \in \mathbb{C}[X]_{\leq n-1}$ and $s \in \mathbb{C}[X]_{\leq m-1}$ such that $T(r, s) = rp + sq = 1$.

□

Chapter 5

§5A

Solution 5A.1

(a) If $U \subseteq \ker T$, then for any $u \in U$, we have $T(u) = 0 \in U$, so U is invariant under T.

(b) If $\operatorname{im} T \subseteq U$, then for any $u \in U$, $T(u) \in \operatorname{im} T \subseteq U$, so U is invariant under T.

□

Solution 5A.2 Let $v \in V_1 + \cdots + V_m$. Then $v = v_1 + \cdots + v_m$ for some $v_i \in V_i$. Since each V_i is invariant under T, we have $T(v_i) \in V_i$ for each i. Therefore,

$$T(v) = T(v_1 + \cdots + v_m) = T(v_1) + \cdots + T(v_m) \in V_1 + \cdots + V_m,$$

so $V_1 + \cdots + V_m$ is invariant under T. □

Solution 5A.3 Let $\{U_i\}_{i \in I}$ be a collection of subspaces of V invariant under T, and let

$$U = \bigcap_{i \in I} U_i.$$

Let $u \in U$. Then $u \in U_i$ for all $i \in I$. Since each U_i is invariant under T, we have $T(u) \in U_i$ for all $i \in I$. Therefore, $T(u) \in U$, so U is invariant under T. □

Solution 5A.4 If $U = \{0\}$, then we are done. Now suppose $U \neq \{0\}$ so that there exists nonzero $v_1 \in U$. Extend v_1 to a basis $v_1, \ldots, v_m$ of V, and let $v \in V$. Define the linear operator $T \in \text{End}(V)$ by $T(v_1) = v$ and $T(v_i) = 0$ for $i = 2, \ldots, m$. Since U is invariant under T, we have $T(v_1) = v \in U$. But v was an arbitrary element of V, so $U = V$. Therefore, $U = \{0\}$ or $U = V$. □

Solution 5A.5 If (x, y) is an eigenvector of T with eigenvalue λ, then we have the system of equations

$$\begin{cases} -3y = \lambda x, \\ x = \lambda y \end{cases}$$

Solving the system, we get $\lambda^2 = -3$ which has no real solutions. Therefore, T has no real eigenvalues. □

Solution 5A.6 If (w, z) is an eigenvector of T with eigenvalue λ, then we have

$$(z, w) = \lambda(w, z),$$

which gives the system

$$\begin{cases} z = \lambda w, \\ w = \lambda z. \end{cases}$$

Solving, we get $\lambda^2 = 1$, so $\lambda = 1$ or $\lambda = -1$. If $\lambda = 1$, then $z = w$, so the eigenvectors are of the form (w, w). If $\lambda = -1$, then $z = -w$, so the eigenvectors are of the form $(w, -w)$. □

Solution 5A.7 If (z_1, z_2, z_3) is an eigenvector of T with eigenvalue λ, then we have

$$(2z_2, 0, 5z_3) = \lambda(z_1, z_2, z_3),$$

which gives the system

$$\begin{cases} 2z_2 = \lambda z_1, \\ 0 = \lambda z_2, \\ 5z_3 = \lambda z_3. \end{cases}$$

Solving, we get the eigenvalues $\lambda = 0$ and $\lambda = 5$. For $\lambda = 0$, the eigenvectors are of the form $(z_1, 0, 0)$. For $\lambda = 5$, the eigenvectors are of the form $(0, 0, z_3)$. □

Solution 5A.8 If v is an eigenvector of P with eigenvalue λ, then $Pv = \lambda v$. Applying P again, we get $P^2v = P(\lambda v) = \lambda Pv = \lambda^2 v$. But $P^2 = P$, so $P^2v = Pv = \lambda v$. Therefore, $\lambda^2 v = \lambda v$, which gives $(\lambda^2 - \lambda)v = 0$. Since $v \neq 0$, we must have $\lambda^2 - \lambda = 0$, so $\lambda(\lambda - 1) = 0$. Hence, $\lambda = 0$ or $\lambda = 1$. □

Solution 5A.9 Let $p \in \mathbb{R}[X]$ such that $p' = \lambda p$ for some $\lambda \in \mathbb{R}$.

- If $\lambda \neq 0$, suppose $\deg p = 0$. Then p is a constant polynomial, and $p' = 0 \neq \lambda p$, so $\deg p \geq 1$. If $\deg p = n \geq 1$, then $\deg p' = n - 1$, so $\deg p' \neq \deg p$, which contradicts $p' = \lambda p$.
- If $\lambda = 0$, then $p' = 0$, which implies that p is a constant polynomial. Therefore, the eigenvectors corresponding to the eigenvalue $\lambda = 0$ are all nonzero constant polynomials. □

Solution 5A.10 Let $p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$ be an eigenvector of T with eigenvalue λ. Then

$$(Tp)(x) = xp'(x) = x(a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3) = a_1x + 2a_2x^2 + 3a_3x^3 + 4a_4x^4.$$

Equating coefficients with $\lambda p(x) = \lambda a_0 + \lambda a_1x + \lambda a_2x^2 + \lambda a_3x^3 + \lambda a_4x^4$ yields the system:

$$\lambda a_0 = 0, \quad \lambda a_1 = a_1, \quad \lambda a_2 = 2a_2, \quad \lambda a_3 = 3a_3, \quad \lambda a_4 = 4a_4.$$

For each $n \in \{0, 1, 2, 3, 4\}$, the equation $\lambda a_n = na_n$ is satisfied with $a_n \neq 0$ if and only if $\lambda = n$. When $\lambda = n$, all other coefficients must be zero, so the eigenvector is a nonzero scalar multiple of x^n. Therefore, the eigenvalues of T are $0, 1, 2, 3, 4$, with corresponding eigenvectors being nonzero scalar multiples of $1, x, x^2, x^3, x^4$, respectively. □

Solution 5A.11 Since V is finite-dimensional, the set of eigenvalues of T is finite. If T has no eigenvalues, then $T - \lambda I$ is invertible for all $\lambda \in \mathbb{F}$, and we can choose any $\delta > 0$. Otherwise, let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of T. Define

$$\delta = \min_{\lambda_j \neq \alpha} |\alpha - \lambda_j|.$$

If α is the only eigenvalue, then we can select any $\delta > 0$ since any λ satisfying $0 < |\alpha - \lambda| < \delta$ will not be equal to α, and hence $T - \lambda I$ will be invertible. Otherwise, we have for any $\lambda \in \mathbb{F}$ such that $0 < |\alpha - \lambda| < \delta$, we have $\lambda \neq \lambda_j$ for all $j = 1, \ldots, m$, so λ is not an eigenvalue of T. Hence, $T - \lambda I$ is invertible for all such λ. □

Solution 5A.12 Let $v = u + w \in V$ be an eigenvector of P with eigenvalue λ. Then $P(v) = P(u + w) = u = \lambda v = \lambda(u + w)$. Comparing the U and W components, we get $u = \lambda u$ and $0 = \lambda w$. If $\lambda \neq 0$, then $w = 0$ and $u = \lambda u$ implies $\lambda = 1$. If $\lambda = 0$, then $u = 0$ and w can be any nonzero vector in W. Therefore, the eigenvalues are $\lambda = 1$ and $\lambda = 0$, with corresponding eigenvectors being the nonzero vectors in U for $\lambda = 1$ and the nonzero vectors in W for $\lambda = 0$. □

Solution 5A.13

(a) Let λ be an eigenvalue of T with eigenvector $v \in V$, so $T(v) = \lambda v$. Then

$$(S^{-1}TS)(S^{-1}v) = S^{-1}T(v) = S^{-1}(\lambda v) = \lambda(S^{-1}v).$$

Thus, $S^{-1}v$ is an eigenvector of $S^{-1}TS$ with eigenvalue λ. This shows that every eigenvalue of T is an eigenvalue of $S^{-1}TS$. Conversely, if λ is an eigenvalue of $S^{-1}TS$ with eigenvector u, then Su is an eigenvector of T with eigenvalue λ. Therefore, T and $S^{-1}TS$ have the same eigenvalues.

(b) The relation between the eigenvectors of T and the eigenvectors of $S^{-1}TS$ is given by $v \mapsto S^{-1}v$. Specifically, if v is an eigenvector of T with eigenvalue λ, then $S^{-1}v$ is an eigenvector of $S^{-1}TS$ with the same eigenvalue λ. Conversely, if u is an eigenvector of $S^{-1}TS$ with eigenvalue λ, then Su is an eigenvector of T with eigenvalue λ.

□

Solution 5A.14 Consider the linear operator on $\mathbb{R}^4$ given by the matrix

$$T = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Letting T act on $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$, we have

$$T(x_1, x_2, x_3, x_4) = (-x_2, x_1, -x_4, x_3).$$

Then any eigenvector $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$ with eigenvalue λ must satisfy

$$(-x_2, x_1, -x_4, x_3) = \lambda(x_1, x_2, x_3, x_4).$$

This yields $\lambda^2 + 1 = 0$, which has no solution in $\mathbb{R}$. □

Solution 5A.15

$$\begin{aligned} \lambda \text{ is an eigenvalue of } T &\iff T - \lambda I \text{ is not invertible} \\ &\iff (T - \lambda I)' \text{ is not invertible} \\ &\iff T' - \lambda I \text{ is not invertible} \\ &\iff \lambda \text{ is an eigenvalue of } T' \end{aligned}$$

□

Solution 5A.16

Axler Note. *See Exercise 6A19 for a different bound on $|\lambda|$.*

Let λ be an eigenvalue of T with eigenvector $u \in V$, so $Tu = \lambda u$. Let $\operatorname{Mat}(T)$ be the matrix of T with respect to the basis $v_1, \ldots, v_n$. Then

$$\lambda u_j = (\operatorname{Mat}(T)u)_j = \sum_{k=1}^{n} \operatorname{Mat}(T)_{j,k} u_k \quad \text{for each } j = 1, \ldots, n.$$

Taking absolute values, we get

$$|\lambda||u_j| = \left| \sum_{k=1}^{n} \operatorname{Mat}(T)_{j,k} u_k \right| \le \sum_{k=1}^{n} |\operatorname{Mat}(T)_{j,k}||u_k| \le n \max\{|\operatorname{Mat}(T)_{j,k}| : 1 \le j, k \le n\} \max\{|u_k| : 1 \le k \le n\}.$$

Choosing j such that $|u_j| = \max\{|u_k| : 1 \le k \le n\}$, we obtain

$$|\lambda| \le n \max\{|\operatorname{Mat}(T)_{j,k}| : 1 \le j, k \le n\}. \qquad \square$$

Solution 5A.17

Axler Note. *See Exercise 3B33 for the definition of $T_{\mathbb{C}}$.*

Suppose λ is an eigenvalue of T with eigenvector $v \in V$, so $Tv = \lambda v$. Then in the complexification, we have $T_{\mathbb{C}}(v + i0) = Tv + iT0 = \lambda v + i0 = \lambda(v + i0)$, showing that λ is an eigenvalue of $T_{\mathbb{C}}$ with eigenvector $v + i0$.

Conversely, suppose λ is an eigenvalue of $T_{\mathbb{C}}$ with eigenvector $v + iw \in V_{\mathbb{C}}$, where $v, w \in V$. Then

$$T_{\mathbb{C}}(v + iw) = Tv + iTw = \lambda(v + iw) = \lambda v + i\lambda w.$$

Equating real and imaginary parts, we get

$$Tv = \lambda v \quad \text{and} \quad Tw = \lambda w.$$

Since $\lambda \in \mathbb{R}$, if $v \neq 0$, then v is an eigenvector of T corresponding to λ. If $v = 0$, then $w \neq 0$ and $Tw = \lambda w$, so w is an eigenvector of T corresponding to λ. This shows that λ is an eigenvalue of T. $\square$

Solution 5A.18 Suppose λ is an eigenvalue of $T_{\mathbb{C}}$ with eigenvector $v + iw \in V_{\mathbb{C}}$, where $v, w \in V$. Then

$$T_{\mathbb{C}}(v + iw) = Tv + iTw = \lambda(v + iw) = \lambda v + i\lambda w.$$

Moreover, we have

$$T_{\mathbb{C}}(v - iw) = Tv - iTw.$$

Note that $v + iw = 0$ if and only if both $v = w = 0$. Then $v + iw \neq 0$ implies $v - iw \neq 0$. Let $\lambda = a + bi$ with $a, b \in \mathbb{R}$. Then equating real and imaginary parts, we have

$$Tv = av - bw \quad \text{and} \quad Tw = bv + aw.$$

Then

$$\begin{aligned} T_{\mathbb{C}}(v - iw) &= Tv - iTw \\ &= (av - bw) - i(bv + aw) \\ &= av - ibv - bw - iaw \\ &= (a - ib)(v - iw) \\ &= \overline{\lambda}(v - iw). \end{aligned}$$

Since $v - iw \neq 0$, this shows that $\overline{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$ with eigenvector $v - iw$. Conversely, if $\overline{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$, then by the same argument, λ is an eigenvalue of $T_{\mathbb{C}}$. This completes the proof. □

Solution 5A.19 Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of T with eigenvector $(z_1, z_2, \ldots) \in \mathbb{F}^\infty$. Then

$$T(z_1, z_2, \ldots) = (0, z_1, z_2, \ldots) = \lambda(z_1, z_2, \ldots).$$

Equating the first component, we get $0 = \lambda z_1$, so either $\lambda = 0$ or $z_1 = 0$. If $\lambda = 0$, then the second component gives $z_1 = 0$, and the third component gives $z_2 = 0$, and so on. By induction, all $z_k = 0$, contradicting that the eigenvector is nonzero. If $\lambda \neq 0$, then $z_1 = 0$, and the second component gives $z_2 = 0$, and so on. Again, all $z_k = 0$, a contradiction. Therefore, T has no eigenvalues. □

Solution 5A.20

(a) Let $\alpha \in \mathbb{F}$. Consider the vector $(1, \alpha, \alpha^2, \ldots) \in \mathbb{F}^\infty$. Then

$$S(1, \alpha, \alpha^2, \ldots) = (\alpha, \alpha^2, \alpha^3, \ldots) = \alpha(1, \alpha, \alpha^2, \ldots),$$

showing that α is an eigenvalue of S with eigenvector $(1, \alpha, \alpha^2, \ldots)$. Since $\alpha \in \mathbb{F}$ was arbitrary, every element of $\mathbb{F}$ is an eigenvalue of S.

(b) Let $\alpha \in \mathbb{F}$ and consider the eigenvalue α of S. Any eigenvector corresponding to α must satisfy

$$S(z_1, z_2, z_3, \ldots) = \alpha(z_1, z_2, z_3, \ldots),$$

which gives

$$(z_2, z_3, z_4, \ldots) = (\alpha z_1, \alpha z_2, \alpha z_3, \ldots).$$

Equating components, we get $z_2 = \alpha z_1$, $z_3 = \alpha z_2 = \alpha^2 z_1$, and in general $z_k = \alpha^{k-1} z_1$ for $k \geq 1$. Therefore, all eigenvectors corresponding to α are of the form $(z_1, \alpha z_1, \alpha^2 z_1, \ldots) = z_1(1, \alpha, \alpha^2, \ldots)$ with $z_1 \neq 0$.

□

Solution 5A.21

(a) Suppose $\lambda \in \mathbb{F}$ with $\lambda \neq 0$ is an eigenvalue of T with eigenvector $v \in V$. Then

$$Tv = \lambda v.$$

Since T is invertible, we can apply T^{-1} to both sides to get

$$v = T^{-1}(\lambda v) = \lambda T^{-1} v.$$

Dividing both sides by λ (which is nonzero), we obtain

$$T^{-1}v = \frac{1}{\lambda}v.$$

This shows that $\frac{1}{\lambda}$ is an eigenvalue of T^{-1} with the same eigenvector v.

Conversely, suppose $\mu \in \mathbb{F}$ with $\mu \neq 0$ is an eigenvalue of T^{-1} with eigenvector $v \in V$. Then

$$T^{-1}v = \mu v.$$

Applying T to both sides gives

$$v = \mu T v \implies Tv = \frac{1}{\mu}v.$$

This shows that $\frac{1}{\mu}$ is an eigenvalue of T with the same eigenvector v.

Therefore, λ is an eigenvalue of T if and only if $\frac{1}{\lambda}$ is an eigenvalue of T^{-1}.

(b) Part (a) shows that if λ is an eigenvalue of T, then $\frac{1}{\lambda}$ is an eigenvalue of T^{-1} with the same eigenvector. Conversely, if μ is an eigenvalue of T^{-1}, then $\frac{1}{\mu}$ is an eigenvalue of T with the same eigenvector. Therefore, T and T^{-1} have the same eigenvectors.

□

Solution 5A.22 Consider the vectors $u+w$ and $u-w$. Since u and w are both nonzero, at least one of $u+w$ or $u-w$ is nonzero. We have

$$T(u+w) = Tu + Tw = 3w + 3u = 3(u+w),$$

and

$$T(u-w) = Tu - Tw = 3w - 3u = -3(u-w).$$

Hence, one of $u+w$ or $u-w$ is an eigenvector with eigenvalue 3 or -3, respectively. □

Solution 5A.23 Let λ be an eigenvalue of ST with eigenvector $v \in V$, so that

$$STv = \lambda v.$$

If $\lambda = 0$, then ST is not invertible. Since V is finite-dimensional, TS is also not invertible, so 0 is an eigenvalue of TS. Otherwise, if $\lambda \neq 0$, let $w = Tv$. Then

$$Sw = S(Tv) = STv = \lambda v \neq 0 \implies w \neq 0.$$

Now, we have

$$TSw = T(STv) = T(\lambda v) = \lambda Tv = \lambda w.$$

Therefore, w is an eigenvector of TS with eigenvalue λ. Then every eigenvalue of ST is also an eigenvalue of TS. By symmetry, the same argument shows that if λ is an eigenvalue of TS, then it is also an eigenvalue of ST. Hence, ST and TS have the same eigenvalues. □

Solution 5A.24

(a) Suppose the sum of the entries in each row of A equals 1. Let $v = (1, 1, \ldots, 1)^t \in \mathbb{F}^n$. Then

$$Tv = Av = v,$$

showing that v is an eigenvector of T corresponding to the eigenvalue 1.

(b) Suppose the sum of the entries in each column of A equals 1. By part (a), this is equivalent to saying that the sum of the entries in each row of A^t equals 1, hence 1 is an eigenvalue of the linear operator defined by A^t. But the eigenvalues of A^t are the same as the eigenvalues of A because $A - \lambda I$ is not invertible if and only if $A^t - \lambda I$ is not invertible. Therefore, 1 is an eigenvalue of T.

□

Solution 5A.25 Let $Tu = \lambda u$ and $Tw = \mu w$ for some scalars $\lambda, \mu \in \mathbb{F}$. Since $u + w$ is also an eigenvector of T, we have $T(u + w) = \nu(u + w)$ for some scalar $\nu \in \mathbb{F}$. Then

$$\lambda u + \mu w = Tu + Tw = T(u + w) = \nu(u + w) = \nu u + \nu w.$$

Suppose u and w correspond to different eigenvalues. Then the list u, w is linearly independent, and we can equate coefficients of u and w above to obtain $\lambda = \nu = \mu$, which is a contradiction. Therefore, u and w must correspond to the same eigenvalue.

□

Solution 5A.26 Let $v \in V$ be any nonzero vector. By assumption, v is an eigenvector of T, so $Tv = \lambda_v v$ for some scalar $\lambda_v \in \mathbb{F}$. Now, let $u \in V$ be another nonzero vector. Consider the vector $u + v$. If u and v are linearly dependent, then without loss of generality, let $u = \alpha v$ for some scalar $\alpha \in \mathbb{F}$. Then $Tu = \lambda_u u = \lambda_u \alpha v$, and $Tu = T(\alpha v) = \alpha Tv = \alpha \lambda_v v$. Equating these, we get $\lambda_u \alpha v = \alpha \lambda_v v$, and since $\alpha \neq 0$ and $v \neq 0$, it follows that $\lambda_u = \lambda_v$.

The remaining case is when $u + v$ are linearly independent. In this case, $u + v$ is an eigenvector of T, so $T(u + v) = \lambda_{u+v}(u + v)$ for some scalar $\lambda_{u+v} \in \mathbb{F}$. On the other hand, $T(u + v) = Tu + Tv = \lambda_u u + \lambda_v v$. Equating these, we get $\lambda_{u+v} u + \lambda_{u+v} v = \lambda_u u + \lambda_v v$. Since u and v are linearly independent, we can equate coefficients to get $\lambda_{u+v} = \lambda_u = \lambda_v$.

Since u and v were arbitrary nonzero vectors, it follows that there exists a scalar $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$ for all nonzero $v \in V$. Hence, $T = \lambda I$. □

Solution 5A.27

Appendix. *An alternate proof is given at the end of this solution.*

Let $v_1 \in V$ be nonzero. Our goal is to intersect enough k-dimensional subspaces to isolate $\operatorname{span} v_1$. We proceed as follows. Let $\dim V = n$, and extend v_1 to a basis $v_1, \ldots, v_n$ of V. For each $i = 2, \ldots, n$, let

$$U_i = \operatorname{span}(v_1, w_2, \ldots, w_k), \quad w_2, \ldots, w_k \in \{v_2, \ldots, v_n\} \setminus \{v_i\},$$

i.e., each U_i must contain v_1 as well as $k - 1$ vectors from the basis except v_i. This choice is possible, since $k \leq n - 1$, so there are enough vectors in $\{v_2, \ldots, v_n\} \setminus \{v_i\}$

to choose $k-1$ of them. The intersection of all such U_i will be exactly span v_1. Since each U_i is invariant under T, we have $Tv_1 \in U_i$ for all i. Therefore, $Tv_1 \in \operatorname{span} v_1$, which means Tv_1 is a scalar multiple of v_1. Since v_1 was arbitrary, it follows that T is a scalar multiple of the identity operator.

Alternate proof. This solution is more elegant than the one provided in the main text, as it avoids the explicit construction of subspaces and instead relies on the intersection of all k-dimensional invariant subspaces containing a given vector.

Alternate proof. Civil Service Pigeon. Let $v \in V$ be nonzero. For any $u \notin \operatorname{span}(v)$, we can extend v, u to a basis and construct a k-dimensional subspace that contains v but not u since $k \in \{1, \dots, \dim V - 1\}$. So, the intersection of all k-dimensional subspaces containing v is precisely $\operatorname{span}(v)$.

The intersection of invariant subspaces is invariant, so $\operatorname{span}(v)$ is invariant. This means that every nonzero vector is an eigenvector of T, meaning that T is a scalar multiple of the identity by Exercise 5A26. □

Solution 5A.28 Let $\lambda_1, \dots, \lambda_m$ be distinct eigenvalues of T with corresponding eigenvectors $v_1, \dots, v_m$. Since the eigenvalues are distinct, at most one eigenvalue must be 0. Without loss of generality, assume $\lambda_1 = 0$ if 0 is an eigenvalue. Then $\lambda_2, \dots, \lambda_m$ are nonzero eigenvalues of T. For $j = 2, \dots, m$, we have $Tv_j = \lambda_j v_j$, so $T(\lambda_j^{-1} v_j) = v_j$, which implies $v_j \in \operatorname{im} T$. Therefore, $v_2, \dots, v_m$ are linearly independent vectors in $\operatorname{im} T$, so $m - 1 \leq \dim \operatorname{im} T$. Hence, $m \leq 1 + \dim \operatorname{im} T$. □

Solution 5A.29 Since $\dim \mathbb{R}^3 = 3$, the eigenvalues of T are exactly $-4, 5$, and $\sqrt{7}$. It follows that 9 is not an eigenvalue of T, and $\ker(T - 9I) = \{0\}$. Then $T - 9I$ is injective, and since $\mathbb{R}^3$ is finite-dimensional, $T - 9I$ is also surjective. Therefore, for any vector in $\mathbb{R}^3$, there exists $x \in \mathbb{R}^3$ such that $(T - 9I)x = (-4, 5, \sqrt{7})$. □

Solution 5A.30 Let v be an eigenvector of T corresponding to the eigenvalue λ. Then

$$(T-2I)(T-3I)(T-4I)v = (T-2I)(T-3I)(\lambda-4)v = (T-2I)(\lambda-3)(\lambda-4)v = (\lambda-2)(\lambda-3)(\lambda-4)v = 0.$$

Since $v \neq 0$, it must be that $(\lambda - 2)(\lambda - 3)(\lambda - 4) = 0$. Therefore, $\lambda = 2$, $\lambda = 3$, or $\lambda = 4$. □

Solution 5A.31 Consider the rotation matrix by $\pi/4$ radians:

$$T = \begin{pmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{pmatrix}.$$

Then

$$T^4 = \begin{pmatrix} \cos(\pi) & -\sin(\pi) \\ \sin(\pi) & \cos(\pi) \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -I. \qquad \square$$

Solution 5A.32

Manual Remark. *The proof of this is a little easier if we may use minimal polynomials.*

Note that

$$T^4 - I = (T^2 - I)(T^2 + I) = (T - I)(T + I)(T^2 + I) = 0.$$

Since T has no eigenvalues, $T-I$ and $T+I$ are both injective, hence T^2-I is injective. Then for any $v \in V$, we have

$$(T^2 - I)(T^2 + I)v = 0,$$

which implies $(T^2 + I)v = 0$ since $T^2 - I$ is injective. Therefore, $T^2 + I = 0$, or $T^2 = -I$. □

Solution 5A.33

(a) Suppose T is injective. Then if $T^m v = 0$, we have $T(T^{m-1}v) = 0$, which implies $T^{m-1}v = 0$ since T is injective. Repeating this argument, we eventually get $v = 0$. Hence, T^m is injective.

Conversely, suppose T^m is injective. If $Tv = 0$, then $T^m v = 0$, which implies $v = 0$. Hence, T is injective.

(b) Suppose T is surjective. Then for any $v \in V$, there exists $w \in V$ such that $Tw = v$. Similarly, there exists $w_2 \in V$ such that $Tw_2 = w$, and so on. Repeating this m times, we find $u \in V$ such that $T^m u = v$. Hence, T^m is surjective.

Conversely, suppose T^m is surjective. Then for any $v \in V$, there exists $u \in V$ such that $T^m u = v$. Let $w = T^{m-1}u$. Then $Tw = T(T^{m-1}u) = T^m u = v$. Hence, T is surjective.

□

Solution 5A.34 Suppose $v_1, \ldots, v_m$ is linearly independent. Since V is finite-dimensional, we can extend this list to a basis $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ of V. Define $T \in \text{End}(V)$ by setting $Tv_i = \lambda_i v_i$ for $i = 1, \ldots, m$, where $\lambda_1, \ldots, \lambda_m$ are distinct scalars, and choose $\lambda_{m+1}, \ldots, \lambda_n$ arbitrarily (distinct from each other and from $\lambda_1, \ldots, \lambda_m$). Then $v_1, \ldots, v_m$ are eigenvectors of T corresponding to distinct eigenvalues.

Conversely, suppose there exists $T \in \text{End}(V)$ such that $v_1, \ldots, v_m$ are eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. By 5.11, eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence, $v_1, \ldots, v_m$ is linearly independent. □

Solution 5A.35

Axler Hint. *Let* $V = \text{span}(e^{\lambda_1 x}, \ldots, e^{\lambda_n x})$*, and define an operator* $D \in \text{End}(V)$ *by* $Df = f'$*. Find eigenvalues and eigenvectors of* D*.*

Suppose there exists a linear combination

$$a_1 e^{\lambda_1 x} + \cdots + a_n e^{\lambda_n x} = 0$$

for all $x \in \mathbb{R}$. Let $V = \operatorname{span}(e^{\lambda_1 x}, \ldots, e^{\lambda_n x})$ and define $D \in \operatorname{End}(V)$ by $Df = f'$. Then $De^{\lambda_i x} = \lambda_i e^{\lambda_i x}$, so $e^{\lambda_i x}$ is an eigenvector of D corresponding to the eigenvalue λ_i. Since the λ_i are distinct, by 5.11, the eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence, $a_1 = \cdots = a_n = 0$, proving that $e^{\lambda_1 x}, \ldots, e^{\lambda_n x}$ are linearly independent. □

Solution 5A.36 Suppose there exists a linear combination

$$a_1 \cos(\lambda_1 x) + \cdots + a_n \cos(\lambda_n x) = 0$$

for all $x \in \mathbb{R}$. Let $V = \operatorname{span}(\cos(\lambda_1 x), \sin(\lambda_1 x), \ldots, \cos(\lambda_n x), \sin(\lambda_n x))$ and define $D \in \operatorname{End}(V)$ by $Df = f'$. Then $D^2 \cos(\lambda_i x) = -\lambda_i^2 \cos(\lambda_i x)$, so $\cos(\lambda_i x)$ is an eigenvector of D^2 corresponding to the eigenvalue $-\lambda_i^2$. Since the λ_i are distinct and positive, the $-\lambda_i^2$ are distinct, and by 5.11, the eigenvectors corresponding to distinct eigenvalues are linearly independent. Hence, $a_1 = \cdots = a_n = 0$, proving that $\cos(\lambda_1 x), \ldots, \cos(\lambda_n x)$ are linearly independent. □

Solution 5A.37 Suppose λ is an eigenvalue of T with corresponding eigenvector $v \in V$, so $Tv = \lambda v$. Consider a nonzero $\phi \in V'$ and consider $S \in \operatorname{End}(V)$ defined by $Sw = \phi(w)v$ for all $w \in V$. Then $S \neq 0$ and for all $w \in V$,

$$TSw = T(\phi(w)v) = \phi(w)Tv = \phi(w)\lambda v = \lambda\phi(w)v = \lambda Sw.$$

Hence, λ is an eigenvalue of $\mathcal{A}$.
Conversely, if λ is an eigenvalue of $\mathcal{A}$ with corresponding eigenvector $S \in \operatorname{End}(V)$, then $\mathcal{A}(S) = TS = \lambda S$. Since $S \neq 0$, there exists $v \in V$ such that $Sv \neq 0$. Then $TSv = \lambda Sv$, so λ is an eigenvalue of T. Therefore, the set of eigenvalues of T equals the set of eigenvalues of $\mathcal{A}$. □

Solution 5A.38

(a) To show that the definition of T/U makes sense, suppose $v + U = v' + U$ for some $v, v' \in V$. Then $v - v' \in U$, and since U is invariant under T, we have $T(v - v') \in U$, so $Tv - Tv' \in U$, which implies $Tv + U = Tv' + U$. Hence, the definition of T/U is well-defined.

To show that T/U is an operator on V/U, note that for any $v + U \in V/U$, $(T/U)(v + U) = Tv + U \in V/U$. Linearity follows from the linearity of T: for any $\alpha, \beta \in \mathbb{F}$ and $v + U, w + U \in V/U$,

$$\begin{aligned}(T/U)(\alpha(v + U) + \beta(w + U)) &= (T/U)(\alpha v + \beta w + U)\\ &= T(\alpha v + \beta w) + U\\ &= \alpha Tv + \beta Tw + U\\ &= \alpha(T/U)(v + U) + \beta(T/U)(w + U)\end{aligned}$$

(b) Let λ be an eigenvalue of T/U with eigenvector $v + U \neq 0$. Then $v \notin U$ and $(T - \lambda I)v \in U$. Consider $(T - \lambda I)|_U \in \operatorname{End}(U)$.

- If $(T - \lambda I)|_U$ is not injective, then λ is an eigenvalue of $T|_U$ and hence of T.

- If $(T-\lambda I)|_U$ is injective, then by finite-dimensionality it is surjective onto U. Thus, there exists $u \in U$ with $(T-\lambda I)u = (T-\lambda I)v$. Consequently, $(T-\lambda I)(v-u) = 0$ with $v-u \neq 0$ (since $v \notin U$ and $u \in U$), showing λ is an eigenvalue of T.

In either case, λ is an eigenvalue of T.

□

Solution 5A.39 Suppose T has an eigenvalue λ with corresponding eigenvector $v \in V$. Consider $\operatorname{im}(T-\lambda I)$. Then $\dim \operatorname{im}(T-\lambda I) \leq \dim V - 1$. Extend $\operatorname{im}(T-\lambda I)$ to a subspace U of V of dimension $\dim V - 1$. Then U is invariant under T because for any $u \in U$, $Tu = (T-\lambda I)u + \lambda u \in U$. Hence, there exists a subspace of V of dimension $\dim V - 1$ that is invariant under T.
Conversely, suppose there exists a subspace U of V of dimension $\dim V - 1$ that is invariant under T. Then V/U is one-dimensional, and the induced operator T/U on V/U must have an eigenvalue λ. By Exercise 5A38 (b), any eigenvalue of T/U is also an eigenvalue of T. Hence, T has an eigenvalue. □

Solution 5A.40 We first prove by induction that $(STS^{-1})^k = ST^kS^{-1}$ for all non-negative integers k. For $k = 0$, we have $(STS^{-1})^0 = I = ST^0S^{-1}$. Suppose the result holds for some $k = n$, so that $(STS^{-1})^n = ST^nS^{-1}$. Then

$$\begin{aligned}(STS^{-1})^{n+1} &= (STS^{-1})^n(STS^{-1}) \\ &= (ST^nS^{-1})(STS^{-1}) \\ &= ST^n(S^{-1}S)TS^{-1} \\ &= ST^nTS^{-1} \\ &= ST^{n+1}S^{-1}.\end{aligned}$$

By induction, the result holds for all non-negative integers k.
Now let $p(z) = a_0 + a_1z + \cdots + a_mz^m$ be any polynomial. Then

$$\begin{aligned}p(STS^{-1}) &= a_0I + a_1(STS^{-1}) + \cdots + a_m(STS^{-1})^m \\ &= a_0I + a_1(STS^{-1}) + \cdots + a_m(ST^mS^{-1}) \\ &= S(a_0I + a_1T + \cdots + a_mT^m)S^{-1} \\ &= Sp(T)S^{-1}.\end{aligned}$$

Hence, the result holds for all polynomials p. □

Solution 5A.41 Let $u \in U$. Since U is invariant under T, we have $T(u) \in U$. By induction, suppose $T^k(u) \in U$ for some non-negative integer k. Then $T^{k+1}(u) = T(T^k(u)) \in U$ since U is invariant under T. Hence, $T^k(u) \in U$ for all non-negative integers k. Now let $p(z) = a_0 + a_1z + \cdots + a_mz^m$ be any polynomial. Then

$$p(T)(u) = a_0u + a_1T(u) + \cdots + a_mT^m(u) \in U,$$

since each term is in U. Therefore, U is invariant under $p(T)$ for every polynomial $p \in \mathbb{F}[X]$. □

Solution 5A.42

Appendix. *An alternate proof is given at the end of this solution.*

(a) Let λ be an eigenvalue of T with corresponding eigenvector $v = (x_1, x_2, \ldots, x_n) \in \mathbb{F}^n$. Then $T(v) = \lambda v$, which means

$$(x_1, 2x_2, 3x_3, \ldots, nx_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n).$$

Comparing components, we get $\lambda x_i = i x_i$ for each $i = 1, 2, \ldots, n$. Hence, if $x_i \neq 0$, then $\lambda = i$. Therefore, the eigenvalues of T are $1, 2, \ldots, n$, and the corresponding eigenvectors are nonzero multiples of the standard basis vectors $e_1, e_2, \ldots, e_n$.

(b) Let U be an invariant subspace of $\mathbb{F}^n$ under T. Let

$$u = \sum_{i=1}^{n} \alpha_i e_i \in U,$$

where e_i are the standard basis vectors of $\mathbb{F}^n$ and $\alpha_i \in \mathbb{F}$. Since U is invariant under T, we have

$$Tu = \sum_{i=1}^{n} \alpha_i T(e_i) = \sum_{i=1}^{n} \alpha_i i e_i \in U.$$

Consider some index i such that $\alpha_i \neq 0$. Then

$$u_1 = Tu - iu = \sum_{j=1}^{n} \alpha_j j e_j - i \sum_{j=1}^{n} \alpha_j e_j = \sum_{j=1}^{n} \alpha_j (j - i) e_j \in U.$$

The i-th term in this sum is $\alpha_i (i - i) e_i = 0$, hence the e_i term is eliminated from u_1. Consider another index j of u_1 such that $\alpha_j \neq 0$. Then we can form

$$u_2 = Tu_1 - ju_1 = \sum_{k=1}^{n} \alpha_k (k-i) k e_k - j \sum_{k=1}^{n} \alpha_k (k-i) e_k = \sum_{k=1}^{n} \alpha_k (k-i)(k-j) e_k \in U.$$

The j-th term in this sum is $\alpha_j (j - i)(j - j) e_j = 0$, and u_2 now has all terms except the i-th and j-th terms. By repeating this process for each index with a nonzero coefficient, we can isolate each $\alpha_i e_i$ in U whenever $\alpha_i \neq 0$. This shows that for any $u \in U$, all the standard basis vectors e_i corresponding to nonzero coefficients in u must also be in U. Therefore, U must be the span of some subset of the standard basis vectors $\{e_1, e_2, \ldots, e_n\}$.

Alternate proof. The solution is meant to generalize to any diagonal operator, showing that the invariant subspaces are precisely the spans of subsets of the standard basis vectors. The original solution is meant to show the student the mechanism by which invariant subspaces of a diagonal operator can be determined using the standard basis vectors.

Alternate proof. Civil Service Pigeon. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$.

(a) For each $k \in \{1, \ldots, n\}$, we have $Te_k = ke_k$. Hence, the eigenvalues of T are 1, 2, …, n and the eigenvectors corresponding to k are the nonzero scalar multiples of e_k. Note that this list of eigenvalues (and eigenvectors) is exhaustive because $\dim(\mathbb{F}^n) = n$.

(b) Let $U \subseteq \mathbb{F}^n$ be an invariant subspace under T. Take any $u = \sum_{i=1}^n \alpha_i e_i \in U$ and suppose $\alpha_k \neq 0$ for some k. By Lagrange interpolation, there exists a polynomial p such that $p(i) = \delta_{i,k}$ for all $i \in \{1, \ldots, n\}$. Since U is invariant under T, it is also invariant under $p(T)$, and so

$$p(T)u = \sum_{i=1}^n \alpha_i \, p(T)e_i = \sum_{i=1}^n \alpha_i \, p(i)e_i = \alpha_k e_k \in U \implies e_k \in U.$$

Repeating this argument for every index with a nonzero coefficient in some element of U shows that U contains each of the basis vectors $e_1, \ldots, e_n$. Hence, $U = \operatorname{span}\{e_i : i \in S\}$ for some $S \subseteq \{1, \ldots, n\}$.

Conversely, any such span is invariant because $Te_i = ie_i$ belongs to that span. There are 2^n such subspaces, and they exhaust all invariant subspaces of T. □

Solution 5A.43

Appendix. *An alternate proof is given at the end of this solution.*

If T is the zero transformation or a scalar multiple of the identity, the result trivially holds. Otherwise, because $\dim V > 1$, there is at least one nonzero vector $v_1 \in V$ that is not an eigenvector of T. Let $Tv_1 = v_2$, so that v_1 and v_2 are linearly independent, and extend v_1, v_2 to a basis $v_1, \ldots, v_n$ for V.

Now suppose that $\{p(T) : p \in \mathbb{F}[X]\} = \operatorname{End}(V)$ and consider the map $S \in \operatorname{End}(V)$ defined by $Sv_1 = 0$ and $Sv_i = v_i$ for $i = 2, \ldots, n$. Because $S \in \operatorname{End}(V)$, there must be a polynomial $p \in \mathbb{F}[X]$ such that $p(T) = S$. Because the null space of $p(T)$ is invariant under T, it must be that the null space of S is invariant under T. This is impossible, though, as $Sv_1 = 0$ and $STv_1 = Sv_2 = v_2 \neq 0$. Therefore, there is no $T \in \operatorname{End}(V)$ such that $\{p(T) : p \in \mathbb{F}[X]\} = \operatorname{End}(V)$.

Alternate proof. This solution uses the commutativity of operators in $\{p(T) : p \in \mathbb{F}[X]\}$ to show that not all operators in $\operatorname{End}(V)$ can be expressed as a polynomial in T.

Alternate proof. Let $v_1, \ldots, v_n$ be a basis for V, with $n = \dim V$. Consider the operators $T_1, T_2 \in \operatorname{End}(V)$ defined by

$$T_1 v_1 = v_2, \quad T_1 v_i = 0 \text{ for } i = 2, \ldots, n,$$

and

$$T_2 v_1 = v_1, \quad T_2 v_i = 0 \text{ for } i = 2, \ldots, n.$$

Then $T_2 T_1 v_1 = T_2(T_1 v_1) = T_2 v_2 = 0$, and $T_1 T_2 v_1 = T_1(T_2 v_1) = T_1 v_1 = v_2$. Hence, $T_2 T_1 \neq T_1 T_2$, so T_1 and T_2 do not commute. Let $S = \{p(T) : p \in \mathbb{F}[X]\}$. Let $p, q \in \mathbb{F}[X]$ be defined as

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m, \quad q(z) = b_0 + b_1 z + \cdots + b_k z^k.$$

Since $p(T)q(T) = q(T)p(T)$, it follows that all operators in S commute. However, if $T_1 = r(T)$ and $T_2 = s(T)$ for some polynomials $r, s \in \mathbb{F}[X]$, then T_1 and T_2 would commute, which is a contradiction. Hence, T_1 and T_2 cannot both be in S, so $S \neq \operatorname{End}(V)$. □

§5B

Solution 5B.1 Suppose 9 is an eigenvalue of T^2. Then there exists a nonzero vector $v \in V$ such that $T^2 v = 9v$. This implies $(T - 3I)(T + 3I)v = 0$. Let $w = (T + 3I)v$, and suppose $w \neq 0$. Then

$$(T - 3I)w = Tw - 3w = 0 \implies Tw = 3w,$$

showing that 3 is an eigenvalue of T. If $w = 0$, then $(T + 3I)v = 0$, which implies $Tv = -3v$, showing that -3 is an eigenvalue of T.
Conversely, suppose 3 is an eigenvalue of T. Then there exists a nonzero vector $v \in V$ such that $Tv = 3v$. Applying T again, we get $T^2 v = T(3v) = 3Tv = 3 \cdot 3v = 9v$, showing that 9 is an eigenvalue of T^2. Similarly, if -3 is an eigenvalue of T, then 9 is an eigenvalue of T^2. □

Solution 5B.2 Suppose U is a subspace of V invariant under T. If $U = \{0\}$, we are done. Suppose U is nonzero and finite-dimensional. By invariance of U under T, $T|_U \in \operatorname{End}(U)$, and $T|_U$ has an eigenvalue $\lambda \in \mathbb{C}$ with corresponding eigenvector $v \in U$, $v \neq 0$. But then v is also an eigenvector of T with eigenvalue λ, contradicting the assumption that T has no eigenvalues. Hence, any nonzero invariant subspace U must be infinite-dimensional. □

Solution 5B.3

(a) If $n = 1$, then $T(x_1) = (x_1)$, and the only eigenvalue is 1 with corresponding eigenvectors being all nonzero vectors in $\mathbb{F}^1$. Otherwise, suppose $T(x_1, \ldots, x_n) = \lambda(x_1, \ldots, x_n)$ for some $\lambda \in \mathbb{F}$ and $(x_1, \ldots, x_n) \neq 0$. Then

$$\lambda x_j = x_1 + \cdots + x_n \quad \text{for all } j = 1, \ldots, n.$$

This implies that $\lambda x_1 = \lambda x_2 = \cdots = \lambda x_n$. If $\lambda \neq 0$, then $x_1 = x_2 = \cdots = x_n$, and we can write $(x_1, \ldots, x_n) = (x_1, \ldots, x_1) = x_1(1, \ldots, 1)$. Substituting this into the eigenvalue equation gives

$$\lambda(x_1, \ldots, x_1) = T(x_1, \ldots, x_1) = (nx_1, \ldots, nx_1),$$

which implies $\lambda = n$. If $\lambda = 0$, then $x_1 + \cdots + x_n = 0$, and the eigenvectors are all nonzero vectors $(x_1, \ldots, x_n)$ such that $x_1 + \cdots + x_n = 0$. Hence, the eigenvalues of T are 0 and n, with corresponding eigenvectors as described.

(b) If $n = 1$, the minimal polynomial is $z - 1$.

For $n > 1$, we know from (a) that the eigenvalues are 0 and n. Then the minimal polynomial is a polynomial multiple of $z(z - n)$. Letting $x = x_1 + \cdots + x_n$, then

$$T(x_1, \ldots, x_n) = (x, \ldots, x), \quad (T - nI)(x_1, \ldots, x_n) = (x - nx_1, \ldots, x - nx_n).$$

Observe that $(x - nx_1) + \cdots + (x - nx_n) = nx - n(x_1 + \cdots + x_n) = 0$. Therefore, $(T - nI)(x_1, \ldots, x_n)$ lies in $\ker T$, and applying T again gives

$$T(T - nI)(x_1, \ldots, x_n) = 0.$$

Hence, $z(z-n)$ is a polynomial multiple of the minimal polynomial, and since it must be monic, the minimal polynomial of T is $z(z-n)$.

□

Solution 5B.4 Suppose α is an eigenvalue of $p(T)$. Then there exists a nonzero vector $v \in V$ such that $p(T)v = \alpha v$. Then $p(T) - \alpha I$ is not injective. If $p(z) - \alpha = 0$, then $\alpha = p(\lambda)$ for any $\lambda \in \mathbb{C}$, and we are done. If $p(z) - \alpha = \beta$ for some constant $\beta \in \mathbb{F}$, then $p(T) - \alpha I = \beta I$ is injective, which contradicts the fact that $p(T) - \alpha I$ is not injective. We may now assume that $p(z) - \alpha$ is a nonconstant polynomial. By the Fundamental Theorem of Algebra, we can factor $p(z) - \alpha$ as

$$p(z) - \alpha = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m)$$

for some $c \in \mathbb{C} \setminus \{0\}$ and $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$. Then

$$p(T) - \alpha I = c(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_m I).$$

Since $p(T) - \alpha I$ is not injective, at least one of the factors on the right-hand side is not injective. Therefore, $T - \lambda_j I$ is not injective for some j, which means λ_j is an eigenvalue of T. Then $p(\lambda_j) - \alpha = 0$, so $\alpha = p(\lambda_j)$ for some eigenvalue λ_j of T. Conversely, suppose $\alpha = p(\lambda)$ for some eigenvalue λ of T with corresponding eigenvector v. Let $p(z) = a_m z^m + \cdots + a_1 z + a_0$. Since $T^m v = \lambda^m v$, we have

$$p(T)v = a_m T^m v + \cdots + a_1 T v + a_0 v = a_m \lambda^m v + \cdots + a_1 \lambda v + a_0 v = p(\lambda)v = \alpha v.$$

Hence, α is an eigenvalue of $p(T)$. □

Solution 5B.5 Consider the operator $T \in \text{End}(\mathbb{R}^2)$ defined by $T(x, y) = (-y, x)$, which represents a counterclockwise rotation by 90°. Observe that T has no real eigenvalues as follows: If (x, y) is an eigenvector of T with eigenvalue $\lambda \in \mathbb{R}$, then $T(x, y) = \lambda(x, y)$ implies $(-y, x) = (\lambda x, \lambda y)$. This gives the system of equations:

$$-y = \lambda x \quad \text{and} \quad x = \lambda y.$$

Since x and y cannot both be zero, one can work out that $\lambda^2 = -1$, which has no real solutions. Therefore, T has no real eigenvalues. Now consider the polynomial $p(z) = z^2 + 1$. Then

$$p(T)(x, y) = T^2(x, y) + (x, y) = (-x, -y) + (x, y) = (0, 0).$$

Thus, 0 is an eigenvalue of $p(T)$ with corresponding eigenvector $(1, 0)$ (or any nonzero vector in $\mathbb{R}^2$). However, there is no real eigenvalue λ of T such that $p(\lambda) = 0$, since the eigenvalues of T are complex. This example demonstrates that the result in Exercise 5B4 does not hold when $\mathbb{C}$ is replaced with $\mathbb{R}$. □

Solution 5B.6 By the computation done in Exercise 5B5, we have $T^2(w, z) = (-w, -z)$, so $T^2 + I = 0$. Therefore, the polynomial $p(z) = z^2 + 1$ annihilates T. To see that this is the minimal polynomial, suppose $T = \alpha I$. Then $T(1, 0) = (0, 1) \neq \alpha(1, 0)$ for any $\alpha \in \mathbb{F}$, so $T \neq \alpha I$. Hence, the minimal polynomial of T is $p(z) = z^2 + 1$. □

Solution 5B.7

Axler Hint. *Show that if S is invertible and $p \in \mathbb{F}[X]$, then $p(TS) = S^{-1}p(ST)S$.*

(a) Consider the operators $S, T \in \text{End}(\mathbb{F}^2)$ defined by $S(w,z) = (0,w)$ and $T(w,z) = (w,0)$. Then

$$ST(w,z) = S(w,0) = (0,w) \quad \text{and} \quad TS(w,z) = T(0,w) = (0,0).$$

The minimal polynomial of ST is $p(z) = z^2$, since $ST \neq 0$ but $(ST)^2 = 0$, while the minimal polynomial of TS is $q(z) = z$. Hence, the minimal polynomials of ST and TS are not equal.

(b) Without loss of generality, suppose S is invertible, and note that $ST = STSS^{-1}$. Let p be the minimal polynomial of TS. By Exercise 5A40, we have

$$p(ST) = p(S(TS)S^{-1}) = Sp(TS)S^{-1} = 0,$$

so p annihilates ST. Similarly, let q be the minimal polynomial of ST. Then

$$q(TS) = q(S^{-1}(ST)S) = S^{-1}q(ST)S = 0$$

as $TS = S^{-1}(ST)S$. Therefore, since p annihilates ST, p divides q, and since q annihilates TS, q divides p, so $p = q$.

□

Solution 5B.8

Axler Note. *Because* $\dim \mathbb{R}^2 = 2$*, the degree of the minimal polynomial of T is at most 2. Thus, the minimal polynomial of T is not the tempting polynomial $x^{180} + 1$, even though $T^{180} = -I$.*

Since T is not a scalar multiple of the identity operator, the minimal polynomial of T cannot have degree 1. Moreover, $\dim \mathbb{R}^2 = 2$ implies that the degree of the minimal polynomial of T must be 2. Therefore, the minimal polynomial of T is of the form

$$p(z) = z^2 + bz + c.$$

Consider the point $(1,0) \in \mathbb{R}^2$. Note that $T(x,y) = (x\cos 1^\circ - y \sin 1^\circ, x \sin 1^\circ + y \cos 1^\circ)$. Then

$$T(1,0) = (\cos 1^\circ, \sin 1^\circ), \quad T^2(1,0) = (\cos 2^\circ, \sin 2^\circ).$$

Since $p(T) = 0$, we have

$$p(T)(1,0) = T^2(1,0) + bT(1,0) + c(1,0) = (\cos 2^\circ + b\cos 1^\circ + c, \sin 2^\circ + b \sin 1^\circ) = (0,0).$$

This gives the system of equations:

$$\begin{cases} \cos 2^\circ + b\cos 1^\circ + c = 0, \\ \sin 2^\circ + b \sin 1^\circ = 0. \end{cases}$$

From this, we get the values of b and c:

$$b = -\frac{\sin 2^\circ}{\sin 1^\circ} = -2\cos 1^\circ, \quad c = -\cos 2^\circ - b\cos 1^\circ = 1.$$

Therefore, the minimal polynomial of T is $p(z) = z^2 - 2\cos 1^\circ z + 1$. □

Solution 5B.9 Let A be the matrix of T with respect to some basis of V. Since all entries of A are rational numbers, A^k also has rational entries for any positive integer k. Then the minimal polynomial $p(z) = z^m + a_{m-1}z^{m-1} + \cdots + a_1 z + a_0$ of T in terms of the matrix A satisfies

$$p(A) = A^m + a_{m-1}A^{m-1} + \cdots + a_1 A + a_0 I = 0.$$

Then this gives a system of $(\dim V)^2$ linear equations in the coefficients $a_0, a_1, \ldots, a_{m-1}$ with rational coefficients. By uniqueness of the minimal polynomial, this system has a unique solution. Since the system has rational coefficients and a unique solution, the solution must also be rational. Therefore, all coefficients of the minimal polynomial of T are rational numbers. □

Solution 5B.10 Let $n = \dim V$ and let $p(z) = z^d + a_{d-1}z^{d-1} + \cdots + a_0$ be the minimal polynomial of T. Then $d \leq n$ and $p(T) = 0$. For any $v \in V$,

$$p(T)v = 0 \implies T^d v = -a_{d-1}T^{d-1}v - \cdots - a_1 Tv - a_0 v \implies T^d v \in \operatorname{span}\left(v, Tv, \ldots, T^{d-1}v\right).$$

Let $U = \operatorname{span}\left(v, Tv, \ldots, T^{d-1}v\right)$.
We claim that U is invariant under T. Take any $u = c_0 v + c_1 Tv + \cdots + c_{d-1}T^{d-1}v \in U$. Then,

$$Tu = c_0 Tv + c_1 T^2 v + \cdots + c_{d-2}T^{d-1}v + c_{d-1}T^d v.$$

Every term on the right-hand side except $T^d v$ is automatically in the spanning set of U. We also showed that $T^d v \in U$ earlier, so $Tu \in U$. Hence, $T(U) \subseteq U$.

Because $v \in U$ and U is invariant under T, we see that $T^m v \in U$ for every $m \geq 0$ (this is immediate by induction). So, for any $m \geq n-1$, we have that

$$\operatorname{span}(v, Tv, \ldots, T^m v) \subseteq U \subseteq \operatorname{span}(v, Tv, \ldots, T^{n-1}v)$$

by using $d - 1 \leq n - 1$ to obtain the second inclusion. Since $\{v, Tv, \ldots, T^{n-1}v\} \subseteq \{v, Tv, \ldots, T^m v\}$, the reverse inclusion is immediate, and so

$$\operatorname{span}(v, Tv, \ldots, T^m v) = \operatorname{span}(v, Tv, \ldots, T^{n-1}v). \qquad \square$$

Solution 5B.11

(a) We compute T^2:

$$T^2 = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 + bc & ac + cd \\ ab + bd & bc + d^2 \end{pmatrix}.$$

Then

$$\begin{aligned} T^2 - (a+d)T + (ad - bc)I &= \begin{pmatrix} a^2 + bc & ac + cd \\ ab + bd & bc + d^2 \end{pmatrix} - (a+d)\begin{pmatrix} a & c \\ b & d \end{pmatrix} + (ad - bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a^2 + bc - (a+d)a + ad - bc & ac + cd - (a+d)c \\ ab + bd - (a+d)b & bc + d^2 - (a+d)d + ad - bc \end{pmatrix} \\ &= \begin{pmatrix} a^2 + bc - a^2 - ad + ad - bc & ac + cd - ac - cd \\ ab + bd - ab - bd & bc + d^2 - d^2 - ad + ad - bc \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

(b) If $b = c = 0$ and $a = d$, then the matrix of T is

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix},$$

which is a scalar multiple of the identity matrix. In this case, the minimal polynomial of T is $z - a$.

Otherwise, the polynomial $z^2 - (a + d)z + (ad - bc)$ annihilates T by part (a). To see that this is the minimal polynomial, we need to check that T is not a scalar multiple of the identity matrix. If T were a scalar multiple of the identity matrix, then we would have $b = c = 0$ and $a = d$, which contradicts our assumption. Therefore, the minimal polynomial of T has degree 2 and is $z^2 - (a + d)z + (ad - bc)$ in this case.

□

Solution 5B.12 For each $k \in \{1, \ldots, n\}$, the standard basis vector e_k satisfies $Te_k = ke_k$, so each k is an eigenvalue. Let $p(z) = (z - 1)(z - 2) \cdots (z - n)$. So,

$$p(T)e_k = p(k)e_k = 0 \cdot e_k = 0 \text{ for all } k \in \{1, \ldots, n\},$$

and since the e_k form a basis of V, $p(T) = 0$. Therefore, the minimal polynomial $m_T(z)$ divides $p(z)$.
Now suppose $m_T(z)$ is the minimal polynomial of T. For each eigenvalue k,

$$0 = m_T(T)e_k = m_T(k)e_k \implies m_T(k) = 0.$$

Thus, $(z - k)$ divides $m_T(z)$ for every $k \in \{1, \ldots, n\}$. Because the factors are distinct, $p(z)$ divides $m_T(z)$.

Since both $p(z)$ and $m_T(z)$ are monic, $p(z) = m_T(z)$. Hence, the minimal polynomial is $p(z) = (z - 1) \cdots (z - n)$. □

Solution 5B.13 Let m be the minimal polynomial of T with degree d. By the Division Algorithm for polynomials, we can write

$$p(z) = q(z)m(z) + r(z),$$

where $q, r \in \mathbb{F}[X]$ and $\deg r < d$. Applying this to T, we have

$$p(T) = q(T)m(T) + r(T).$$

Since $m(T) = 0$, it follows that

$$p(T) = r(T).$$

Thus, there exists a polynomial r such that $p(T) = r(T)$ and $\deg r < d$.
To show uniqueness, suppose there exists another polynomial $s \in \mathbb{F}[X]$ such that $p(T) = s(T)$ and $\deg s < d$. Then we have

$$r(T) = p(T) = s(T).$$

This implies that $(r - s)(T) = 0$. Since $\deg(r - s) < d$, it follows that $r - s$ must be the zero polynomial (otherwise, it would contradict the minimality of m). Hence, $r = s$, proving uniqueness. □

Solution 5B.14 Let $p(z) = 4 + 5z - 6z^2 - 7z^3 + 2z^4 + z^5$ be the minimal polynomial of T. We want to find the minimal polynomial of T^{-1}. Since $p(z)$ is the minimal polynomial of T, we have

$$p(T) = 4I + 5T - 6T^2 - 7T^3 + 2T^4 + T^5 = 0.$$

To find the minimal polynomial of T^{-1}, apply T^{-1} to both sides of the equation until we express it in terms of T^{-1}:

$$4T^{-5} + 5T^{-4} - 6T^{-3} - 7T^{-2} + 2T^{-1} + I = 0.$$

Then the polynomial $q(z) = 4z^5 + 5z^4 - 6z^3 - 7z^2 + 2z + 1$ annihilates T^{-1}. To make it monic, we can divide by the leading coefficient:

$$q(z) = z^5 + \frac{5}{4}z^4 - \frac{3}{2}z^3 - \frac{7}{4}z^2 + \frac{1}{2}z + \frac{1}{4}.$$

Suppose T^{-1} has a minimal polynomial $m(z)$ with $\deg m < 5$, hence $m(T^{-1}) = 0$. Applying $T^{\deg m}$ to both sides gives $n(T) = 0$ for some polynomial $n(z)$ of degree less than 5, which contradicts the minimality of $p(z)$. Therefore, the minimal polynomial of T^{-1} is $q(z)$. □

Solution 5B.15 Since V is a finite-dimensional complex vector space, the operator T has at least one eigenvalue $\lambda_0 \in \mathbb{C}$. By Exercise 5A11, there exists $\delta > 0$ such that $T - \lambda I$ is invertible for all $\lambda \in \mathbb{C}$ with $0 < |\lambda - \lambda_0| < \delta$. This implies that for all such λ, we have $\dim \operatorname{im}(T - \lambda I) = \dim V$. However, at the eigenvalue λ_0, we have $\dim \operatorname{im}(T - \lambda_0 I) < \dim V$ since $\dim \ker(T - \lambda_0 I) > 0$. Therefore,

$$\lim_{\lambda \to \lambda_0} f(\lambda) \neq f(\lambda_0),$$

which shows that f is not continuous at λ_0. □

Solution 5B.16 Let $p(z) = a_0 + a_1 z + \cdots + a_{n-1}z^{n-1} + z^n$. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. We will show that $p(T) = 0$ and that no polynomial of lower degree annihilates T. First, we compute $T^k e_1$ for $k = 0, 1, \ldots, n$:

$$T^0 e_1 = e_1, \quad T^1 e_1 = e_2, \quad T^2 e_1 = e_3, \quad \ldots, \quad T^{n-1} e_1 = e_n, \quad T^n e_1 = -a_0 e_1 - a_1 e_2 - \cdots - a_{n-1} e_n.$$

Then we have

$$\begin{aligned} p(T)e_1 &= a_0 T^0 e_1 + a_1 T^1 e_1 + \cdots + a_{n-1} T^{n-1} e_1 + T^n e_1 \\ &= a_0 e_1 + a_1 e_2 + \cdots + a_{n-1} e_n - a_0 e_1 - a_1 e_2 - \cdots - a_{n-1} e_n \\ &= 0. \end{aligned}$$

Since $T^{j-1} e_1 = e_j$ for $j = 1, 2, \ldots, n$, and $p(T)$ commutes with T, we have

$$p(T)e_j = p(T)T^{j-1} e_1 = T^{j-1} p(T) e_1 = T^{j-1} 0 = 0$$

for all $j = 1, 2, \ldots, n$. Since $\{e_1, e_2, \ldots, e_n\}$ is a basis for $\mathbb{F}^n$, it follows that $p(T) = 0$.

To show that p is the minimal polynomial, suppose there exists a polynomial $q(z) = b_0 + b_1 z + \cdots + b_{m-1} z^{m-1} + z^m$ of degree $m < n$ such that $q(T) = 0$. Then we would have

$$q(T)e_1 = b_0 e_1 + b_1 e_2 + \cdots + b_{m-1} e_m + T^m e_1 = 0.$$

However, $T^m e_1 = e_{m+1}$, which is not in the span of $\{e_1, e_2, \ldots, e_m\}$ since $m < n$. This leads to a contradiction, as e_{m+1} cannot be expressed as a linear combination of $e_1, e_2, \ldots, e_m$. Therefore, no polynomial of degree less than n annihilates T, and the minimal polynomial of T is indeed $p(z) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n$. □

Solution 5B.17 Let $p(z)$ be the minimal polynomial of T. Then we have

$$p(T) = 0.$$

Consider the polynomial $q(z) = p(z + \lambda)$. Then we have

$$q(T - \lambda I) = p((T - \lambda I) + \lambda I) = p(T) = 0.$$

Thus, q annihilates $T - \lambda I$.
To show that q is the minimal polynomial, suppose there exists a polynomial $r(z)$ of degree less than $\deg q$ such that $r(T - \lambda I) = 0$. Then we can write

$$s(z) = r(z - \lambda)$$

for some polynomial $s(z)$ of degree less than $\deg p$. Then we have

$$s(T) = r(T - \lambda I) = 0.$$

However, this contradicts the minimality of p, since $\deg s < \deg p$ and $s(T) = 0$. Therefore, no polynomial of degree less than $\deg q$ annihilates $T - \lambda I$, and the minimal polynomial of $T - \lambda I$ is indeed $q(z) = p(z + \lambda)$. □

Solution 5B.18 Let $p(z)$ be the minimal polynomial of T. Then we have

$$p(T) = 0.$$

Consider the polynomial $q(z) = \lambda^{\deg p} p(\frac{z}{\lambda})$. Then we have

$$q(\lambda T) = \lambda^{\deg p} p\left(\frac{\lambda T}{\lambda}\right) = \lambda^{\deg p} p(T) = 0.$$

Moreover, q has leading term $\lambda^{\deg p}(\frac{z}{\lambda})^{\deg p} = z^{\deg p}$, so q is monic and has the same degree as p. To show that q is the minimal polynomial, suppose there exists a polynomial $r(z)$ of degree less than $\deg q$ such that $r(\lambda T) = 0$. Define $s(z) = r(\lambda z)$. Then we have

$$s(T) = r(\lambda T) = 0.$$

However, $\deg s = \deg r < \deg q = \deg p$, which contradicts the minimality of p. Therefore, no polynomial of degree less than $\deg q$ annihilates λT, and the minimal polynomial of λT is indeed $q(z) = \lambda^{\deg p} p(\frac{z}{\lambda})$. □

Solution 5B.19 Let $p(z)$ be the minimal polynomial of T with degree d. Let $\mathcal{I} = \{I, T, T^2, \ldots, T^{d-1}\}$. We will show that $\mathcal{I}$ is a basis for $\mathcal{E}$.
First, we show that $\dim \mathcal{E} \leq d$. By the Division Algorithm for polynomials, for any polynomial $q \in \mathbb{F}[X]$, we can write

$$q(z) = s(z)p(z) + r(z),$$

where $s, r \in \mathbb{F}[X]$ and $\deg r < d$. Then we have

$$q(T) = s(T)p(T) + r(T) = 0 + r(T) = r(T).$$

Since $\deg r < d$, the set $\mathcal{I}$ spans $\mathcal{E}$. Hence, $\dim \mathcal{E} \leq d$.

Next, we show that $\dim \mathcal{E} \geq d$. Suppose

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{d-1} T^{d-1} = 0$$

for scalars $a_0, a_1, \ldots, a_{d-1} \in \mathbb{F}$ not all zero. Then the polynomial

$$q(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{d-1} z^{d-1}$$

annihilates T, contradicting the minimality of $p(z)$ since $\deg q < d$. Therefore, the set $\mathcal{I}$ is linearly independent in $\mathcal{E}$, which implies that $\dim \mathcal{E} \geq d$.

Combining both inequalities, we conclude that $\dim \mathcal{E} = d$, which is the degree of the minimal polynomial of T. □

Solution 5B.20 Let p be the minimal polynomial of T. The eigenvalues of T are 3, 5, and 8, so each is a root of p. Hence,

$$p(z) = (z-3)^{m_1}(z-5)^{m_2}(z-8)^{m_3}$$

where $m_1, m_2, m_3 \geq 1$. Because T acts on a 4-dimensional space, $\deg p \leq 4$, and so $m_1 + m_2 + m_3 \leq 4$. Since $m_2, m_3 \geq 1$, we have $m_1 \leq 4 - 1 - 1 = 2$. By similar logic, $m_1, m_2, m_3 \leq 2$. This means $p(z)$ divides $(z-3)^2(z-5)^2(z-8)^2$. Applying this polynomial to T yields

$$(T - 3I)^2(T - 5I)^2(T - 8I)^2 = 0.$$ □

Solution 5B.21

Axler Note. *If* $\dim \operatorname{im} T < \dim V - 1$, *then this exercise gives a better upper bound than 5.22 for the degree of the minimal polynomial of* T.

Let $r = \dim \operatorname{im} T$. Observe that $\operatorname{im} T$ is invariant under T, and we can consider the restriction of T to $\operatorname{im} T$, denoted by $T|_{\operatorname{im} T}$. Let $p(z)$ be the minimal polynomial of T and let $q(z)$ be the minimal polynomial of $T|_{\operatorname{im} T}$. Then $\deg q \leq r$ since $T|_{\operatorname{im} T}$ is an operator on a space of dimension r.

We now show that the polynomial $zq(z)$ annihilates T. For any $v \in V$, we know that $Tv \in \operatorname{im} T$. By invariance, we know that $w \in \operatorname{im} T$ implies $(T|_{\operatorname{im} T})^k w \in \operatorname{im} T$ for any $k \geq 0$. Then

$$q(T)w = q(T|_{\operatorname{im} T})w = 0.$$

Therefore, for any $v \in V$, we have

$$zq(z)(T)v = Tq(T)v = q(T)Tv = 0,$$

where we used the fact that T commutes with $q(T)$ since $q(T)$ is a polynomial in T. This shows that $zq(z)$ annihilates T, and hence the minimal polynomial $p(z)$ of T divides $zq(z)$. Since $\deg q \leq r$, we have $\deg p \leq 1 + r = 1 + \dim \operatorname{im} T$. □

Solution 5B.22 Suppose T is invertible. Let $q(z)$ be the minimal polynomial of T. Since T is invertible, 0 is not an eigenvalue of T, and thus $q(0) \neq 0$. We may write $q(z)$ as

$$q(z) = z^m + a_{m-1}z^{m-1} + \cdots + a_1 z + a_0,$$

where $a_0 \neq 0$ and $m \leq \dim V$. Then we have

$$0 = q(T) = T^m + a_{m-1}T^{m-1} + \cdots + a_1 T + a_0 I.$$

Rearranging gives

$$I = -\frac{1}{a_0}(T^m + a_{m-1}T^{m-1} + \cdots + a_1 T) \in \operatorname{span}(T, T^2, \ldots, T^m) \subseteq \operatorname{span}(T, T^2, \ldots, T^{\dim V}).$$

Hence, if T is invertible, then $I \in \operatorname{span}(T, T^2, \ldots, T^{\dim V})$.
Conversely, suppose $I \in \operatorname{span}(T, T^2, \ldots, T^{\dim V})$. Then there exist scalars $a_1, a_2, \ldots, a_{\dim V}$ such that

$$I = a_1 T + a_2 T^2 + \cdots + a_{\dim V} T^{\dim V}.$$

Rearranging gives

$$0 = -I + a_1 T + a_2 T^2 + \cdots + a_{\dim V} T^{\dim V} = p(T),$$

where $p(z) = -1 + a_1 z + a_2 z^2 + \cdots + a_{\dim V} z^{\dim V}$. Since the minimal polynomial of T divides any polynomial that annihilates T, the minimal polynomial of T must also have a nonzero constant term. Therefore, 0 is not an eigenvalue of T, and hence T is invertible. □

Solution 5B.23 Let $W = \operatorname{span}(v, Tv, \ldots, T^{n-1}v)$. To show that W is invariant under T, we may show that $Tv, T^2v, \ldots, T^n v \in W$. Observe that $Tv, T(Tv) = T^2v, T(T^2v) = T^3v, \ldots, T(T^{n-2}v) = T^{n-1}v$ are all in W by definition. It remains to show that $T^n v \in W$.
Let $p(z)$ be the minimal polynomial of T. Let $m = \deg p$. Since $m \leq n$, we have

$$p(T)v = T^m v + a_{m-1}T^{m-1}v + \cdots + a_1 Tv + a_0 v = 0.$$

Rearranging gives

$$T^m v = -a_{m-1}T^{m-1}v - \cdots - a_1 Tv - a_0 v \in W.$$

If $m = n$, then we are done. If $m < n$, then we can apply T^{n-m} to both sides to get

$$T^n v = -a_{m-1}T^{n-1}v - \cdots - a_1 T^{n-m+1}v - a_0 T^{n-m}v \in W,$$

where we used the fact that $T^{n-m}v, T^{n-m+1}v, \ldots, T^{n-1}v \in W$ since $n - m < n$. Therefore, $T^n v \in W$, and hence W is invariant under T. □

Solution 5B.24 Let p be the minimal polynomial of T. The eigenvalues of T are 5 and 6, so each is a root of p. Hence,

$$p(z) = (z-5)^{m_1}(z-6)^{m_2}$$

where $m_1, m_2 \geq 1$. Because $\deg p \leq \dim V$, we have $m_1 + m_2 \leq \dim V$. Since $m_2 \geq 1$, we get $m_1 \leq \dim V - 1$, and similarly $m_2 \leq \dim V - 1$. This means $p(z)$ divides $(z-5)^{\dim V - 1}(z-6)^{\dim V - 1}$, so

$$(T-5I)^{\dim V - 1}(T-6I)^{\dim V - 1} = 0. \qquad \square$$

Solution 5B.25

Axler Note. *The quotient operator T/U was defined in Exercise 5A38.*

(a) Let $p(z)$ be the minimal polynomial of T and let $q(z)$ be the minimal polynomial of T/U. Recall that T/U is defined by $(T/U)(v+U) = Tv + U$ for all $v \in V$. Inductively, we have $(T/U)^k(v+U) = T^k v + U$ for all $k \geq 0$. Then we have

$$p(T/U)(v+U) = p(T)v + U = 0 + U = U,$$

where we used the fact that $p(T) = 0$. Therefore, $p(T/U) = 0$, which implies that $q(z)$ divides $p(z)$. Hence, the minimal polynomial of T is a polynomial multiple of the minimal polynomial of the quotient operator T/U.

(b) Let $p(z)$ be the minimal polynomial of T, let $q(z)$ be the minimal polynomial of $T|_U$, and let $r(z)$ be the minimal polynomial of T/U. By definition,

$$r(T/U)(v+U) = r(T)v + U = U,$$

where we used the fact that $r(T/U) = 0$. Then $r(T)v \in U$ for all $v \in V$. Moreover, invariance of U under T implies that $q(T)u = q(T|_U)u = 0$ for all $u \in U$. Since $r(T)v \in U$, and $q(T) = q(T|_U)$ on U, we have

$$q(T)r(T)v = q(T|_U)r(T)v = 0$$

for all $v \in V$. Therefore, $q(T)r(T) = 0$, which implies $p(z)$ divides $q(z)r(z)$. Hence, the product of the minimal polynomials of $T|_U$ and T/U is a polynomial multiple of the minimal polynomial of T.

$\square$

Solution 5B.26 Let λ be an eigenvalue of T with corresponding eigenvector $v \in V \setminus \{0\}$. Then we have

$$Tv = \lambda v.$$

If $v \in U$, then λ is an eigenvalue of $T|_U$. If $v \notin U$, then consider the coset $v + U \in V/U$. We have

$$(T/U)(v+U) = Tv + U = \lambda v + U = \lambda(v+U),$$

which shows that λ is an eigenvalue of T/U.

Conversely, suppose λ is an eigenvalue of $T|_U$ with corresponding eigenvector $u \in U \setminus \{0\}$. Then we have

$$T|_U u = Tu = \lambda u,$$

which shows that λ is also an eigenvalue of T.

Now suppose λ is an eigenvalue of T/U with corresponding eigenvector $v + U \in V/U \setminus \{0\}$. From Exercise 5B25 (a), we know that the minimal polynomial of T is a multiple of the minimal polynomial of T/U. Since λ is an eigenvalue of T/U, it must be a root of the minimal polynomial of T/U. Therefore, λ is also a root of the minimal polynomial of T, which implies that λ is an eigenvalue of T. □

Solution 5B.27

Axler Note. *The complexification $T_{\mathbb{C}}$ was defined in Exercise 3B33.*

Let $p(z)$ be the minimal polynomial of T. Then we have

$$p(T) = 0.$$

Consider the complexification $T_{\mathbb{C}}$ of T. By definition, $T_{\mathbb{C}}$ applied on $v + iw \in V_{\mathbb{C}}$ is given by

$$T_{\mathbb{C}}(v + iw) = T(v) + iT(w)$$

for all $v, w \in V$. By induction, we can extend the polynomial $p(z)$ to act on $V_{\mathbb{C}}$ by defining

$$p(T_{\mathbb{C}})(v + iw) = p(T)v + ip(T)w.$$

Since $p(T) = 0$, it follows that

$$p(T_{\mathbb{C}})(v + iw) = 0 + i0 = 0$$

for all $v, w \in V$. Therefore, $p(T_{\mathbb{C}}) = 0$, which shows that the minimal polynomial of $T_{\mathbb{C}}$ divides $p(z)$.

Conversely, let $q(z)$ be the minimal polynomial of $T_{\mathbb{C}}$. We can split $q(z)$ into its real and imaginary parts, say $q(z) = r(z) + is(z)$, where $r(z), s(z) \in \mathbb{R}[X]$. Then for any $v \in V$ which can be viewed as $v + i0 \in V_{\mathbb{C}}$, we have

$$q(T_{\mathbb{C}})(v + i0) = r(T)(v + i0) + is(T)(v + i0) = r(T)v + is(T)v.$$

Since $q(T_{\mathbb{C}}) = 0$, it follows that $r(T)v = 0$ and $s(T)v = 0$ for all $v \in V$. Therefore, both $r(T)$ and $s(T)$ must be 0, which implies that the minimal polynomial of T divides $r(z)$ and $s(z)$. Since $q(z) = r(z) + is(z)$, it follows that the minimal polynomial of T divides $q(z)$. Combining both directions, we conclude that the minimal polynomial of $T_{\mathbb{C}}$ equals the minimal polynomial of T. □

Solution 5B.28

Axler Note. *The dual map T' was defined in Section 3F, 3.118.*

Let p be the minimal polynomial of T. For any $\phi \in V'$, we have

$$p(T')(\phi) = \phi \circ p(T) = \phi \circ 0 = 0 \implies p(T') = 0.$$

Hence, the minimal polynomial of T' divides p.

Conversely, let q be the minimal polynomial of T', so $q(T') = 0$. Then for all $\phi \in V'$ and all $v \in V$,

$$0 = \big(q(T')\phi\big)(v) = \phi\left(q(T)v\right).$$

Thus, for each fixed v, $\phi(q(T)v) = 0$ for all $\phi \in V'$. Since $\phi(w) = 0$ for all $\phi \in V'$ implies $w = 0$, this forces $q(T)v = 0$ for all v; i.e., $q(T) = 0$. Therefore, p divides q.

Both p and q divide each other and are monic, so they are equal. □

Solution 5B.29

Axler Note. *Exercise 5C6 will give an improvement of this result when* $\mathbb{F} = \mathbb{C}$.

Appendix. *An alternate proof is given at the end of this solution.*

Let V be a finite-dimensional vector space with $\dim V \geq 2$, and let $T \in \text{End}(V)$. Let $p(z)$ be the minimal polynomial of T. Over $\mathbb{F}$, every nonconstant polynomial factors into linear and irreducible quadratic factors. We split the proof into two cases based on the factorization of $p(z)$.

1. Suppose $p(z)$ has a linear factor $z - \lambda$ for some $\lambda \in \mathbb{F}$. If $\deg p(z) = 1$, then $p(z) = z - \lambda$, which implies $T - \lambda I = 0$, or simply $T = \lambda I$. Since T is just a scalar multiple of the identity, every single subspace of V is invariant under T. Because $\dim V \geq 2$, we can choose any two linearly independent vectors $v, w \in V$, and $\text{span}(v, w)$ will form a 2-dimensional invariant subspace.

   If $\deg p(z) \geq 2$, then $p(z)/(z - \lambda)$ has degree at least 1 and must itself contain either a linear factor or an irreducible quadratic factor. The latter is handled by Case 2 below, so we consider the two linear cases:

   - Repeated linear factor: $p(z) = (z - \lambda)^2 q(z)$. Since $p(z)$ is the minimal polynomial, there exists $v \in V$ such that $(T - \lambda I)q(T)v \neq 0$ but $(T - \lambda I)^2 q(T)v = 0$. Let $w = q(T)v$ and $u = (T - \lambda I)w$. Then $u \neq 0$ and $(T - \lambda I)u = (T - \lambda I)^2 w = 0$, which implies $Tu = \lambda u$. Thus, u is an eigenvector corresponding to the eigenvalue λ. Moreover, $Tw = (T - \lambda I)w + \lambda w = u + \lambda w$, which shows that $Tw \in \text{span}(u, w)$.

     To confirm linear independence, suppose $u = \alpha w$ for some $\alpha \in \mathbb{F}$. Then $(T - \lambda I)w = \alpha w$, so w is an eigenvector of T with eigenvalue $\lambda + \alpha$. But then $(T - \lambda I)^2 w = \alpha^2 w$. Since $(T - \lambda I)^2 w = 0$ and $w \neq 0$, this forces $\alpha = 0$, contradicting our deduction that $\alpha \neq 0$ (since $u \neq 0$). Therefore, u and w are linearly independent, and $\text{span}(u, w)$ is a 2-dimensional invariant subspace of V under T.
   - Distinct linear factor: $p(z) = (z - \lambda)(z - \mu)g(z)$ with $\lambda \neq \mu$. Then T has two distinct eigenvalues, yielding two linearly independent eigenvectors v and u. The subspace $\text{span}(v, u)$ is 2-dimensional, and since $Tv = \lambda v \in \text{span}(v, u)$ and $Tu = \mu u \in \text{span}(v, u)$, it is explicitly invariant.

2. Suppose $p(z)$ has an irreducible quadratic factor, say $z^2 + az + b$ for some $a, b \in \mathbb{F}$. This means we can write $p(z) = (z^2 + az + b)q(z)$ for some polynomial $q(z)$. Since $p(z)$ is the minimal polynomial, $q(T)$ is not the zero operator, meaning there exists some vector $x \in V$ such that $q(T)x \neq 0$.

Let $u_1 = q(T)x$. Evaluating the minimal polynomial identity gives:

$$0 = p(T)x = (T^2 + aT + bI)q(T)x = (T^2 + aT + bI)u_1.$$

Rearranging this provides the operator relation $T^2u_1 = -bu_1 - au_2$, where we define $u_2 = Tu_1$.

The vectors u_1 and u_2 must be linearly independent. If $Tu_1 = cu_1$ for some $c \in \mathbb{F}$, then $(T^2 + aT + bI)u_1 = (c^2 + ac + b)u_1 = 0$. Since $u_1 \neq 0$, this implies $c^2 + ac + b = 0$, meaning c is a root of $z^2 + az + b$ over $\mathbb{F}$, contradicting its irreducibility.

Let $U = \operatorname{span}(u_1, u_2)$, which has dimension two. Checking the images of our basis elements under T yields:

$$T(u_1) = u_2 \in U$$
$$T(u_2) = T^2u_1 = -bu_1 - au_2 \in U.$$

Since T maps the generating elements of U back into U, the subspace U is a 2-dimensional invariant subspace of V under T.

Alternate proof. This solution uses upper-triangular matrices, which have not appeared in 5B yet, but they provide a convenient way to identify invariant subspaces of dimension two.

Alternate proof. Civil Service Pigeon. If $\mathbb{F} = \mathbb{C}$, the result is immediate: every operator on a finite-dimensional complex vector space has an upper-triangular matrix with respect to some basis $v_1, \ldots, v_n$. Then $\operatorname{span}(v_1, v_2)$ is a 2-dimensional invariant subspace. Hence, we restrict to $\mathbb{F} = \mathbb{R}$ from now on.

Assume $\dim V \geq 2$ and let $T \in \mathcal{L}(V)$ have minimal polynomial p. Over $\mathbb{R}$, p factors into linear and irreducible quadratic factors.

Case 1: p has an irreducible quadratic factor. Write $p(z) = (z^2+az+b)m(z)$. Pick $v \in V$ such that $x = m(T)v \neq 0$; then $(T^2 + aT + bI)x = 0$. Set $U = \operatorname{span}(x, Tx)$. Clearly, T maps U to itself since $T(Tx) = -bx - aTx \in U$. If x and Tx were linearly dependent, x would be an eigenvector and its eigenvalue would be a real root of $z^2 + az + b$, contradicting irreducibility. Hence $\dim U = 2$.

Case 2: p splits completely into linear factors. Then T is upper-triangularizable over $\mathbb{R}$. Let $v_1, \ldots, v_n$ be a basis with respect to which the matrix of T is upper-triangular. Then, $Tv_1 \in \operatorname{span}(v_1)$ and $Tv_2 \in \operatorname{span}(v_1, v_2)$. Therefore $\operatorname{span}(v_1, v_2)$ is a 2-dimensional invariant subspace.

These two cases cover all possibilities over $\mathbb{R}$, and so combining with trivial complex case finishes the proof. □

§5C

Solution 5C.1

Manual Remark. *The matrix in this exercise represents the* 90° *counterclockwise rotation of* $\mathbb{R}^2$.

We claim that the statement is false. Consider the linear operator $T \in \mathrm{End}(\mathbb{R}^2)$ with matrix representation

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

with respect to the standard basis. Then

$$A^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

which is an upper-triangular matrix with respect to the standard basis of $\mathbb{R}^2$. However, T does not have an upper-triangular matrix with respect to any basis of $\mathbb{R}^2$, since it has no real eigenvalues and thus cannot be represented in upper-triangular form over $\mathbb{R}$. □

Solution 5C.2

(a) Let $C = A + B$. Since A and B are upper-triangular, for any entry c_{ij} of C where $i > j$, we have

$$c_{ij} = a_{ij} + b_{ij} = 0 + 0 = 0,$$

which shows that C is upper-triangular. The diagonal entries of C are given by

$$c_{ii} = a_{ii} + b_{ii} = \alpha_i + \beta_i,$$

for $i = 1, \ldots, n$.

(b) Let $D = AB$. Since A and B are upper-triangular, for any entry d_{ij} of D where $i > j$, we have the following sum for the (i, j)-th entry of D:

$$d_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Consider a_{ik} when $k < i$. Since A is upper-triangular, $a_{ik} = 0$. Now consider b_{kj} when $k > j$. Since B is upper-triangular, $b_{kj} = 0$. Therefore, for $i > j$, all terms in the sum are zero, and we have $d_{ij} = 0$ for $i > j$, showing that D is upper-triangular.

To see that the diagonal entries of D are $\alpha_1\beta_1, \ldots, \alpha_n\beta_n$, we compute the diagonal entries:

$$d_{ii} = \sum_{k=1}^{n} a_{ik} b_{ki},$$

for the i-th diagonal entry. Since A and B are upper-triangular, then $k < i$ implies $a_{ik} = 0$ and $k > i$ implies $b_{ki} = 0$. Therefore, the only nonzero term in the sum occurs when $k = i$, giving

$$d_{ii} = a_{ii} b_{ii} = \alpha_i \beta_i,$$

for $i = 1, \ldots, n$.

□

Solution 5C.3

Appendix. *An alternate proof is given at the end of this solution.*

Let $V_k = \text{span}(v_1, \ldots, v_k)$ for each $k \in \{1, \ldots, n\}$. Because $\text{Mat}(T, (v_1, \ldots, v_n))$ is upper-triangular, $T(V_k) \subseteq V_k$ for every k; thus each V_k is invariant under T. Since T is invertible, its restriction to the (finite-dimensional) V_k is injective, hence surjective onto V_k. Therefore, $T(V_k) = V_k$. Applying T^{-1} yields $V_k = T^{-1}(V_k)$, meaning each V_k is also invariant under T^{-1}. So, $\text{Mat}(T^{-1}, (v_1, \ldots, v_n))$ is upper-triangular.

Let $\mu_1, \ldots, \mu_n$ be the diagonal entries of $\text{Mat}(T^{-1}, (v_1, \ldots, v_n))$. Since $TT^{-1} = I$ is upper-triangular, we have from Exercise 5C2 (b) that $\lambda_k \mu_k = 1$ for each k, giving $\mu_k = \frac{1}{\lambda_k}$.

Alternate proof. This solution attempts for a direct computation of the entries of T^{-1}.

Alternate proof. Let A be the matrix of T with respect to the basis $v_1, \ldots, v_n$. Since T is invertible, A is also invertible. Let $A = (a_{ij})$ be the upper-triangular matrix with $\lambda_1, \ldots, \lambda_n$ on the diagonal, and let $B = (b_{ij})$ be the matrix of T^{-1} with respect to the same basis. Denote b_j as the j-th column of B. Then we have

$$Ab_j = e_j,$$

where e_j is the j-th column of the identity matrix. In the nth row, we have $\lambda_n b_{nj} = \delta_{nj}$, hence $b_{nj} = 0$ for $j < n$ and $b_{nn} = \lambda_n^{-1}$. Now, consider the $(n-1)$-th row. We have

$$a_{n-1,n-1} b_{n-1,j} + a_{n-1,n} b_{n,j} = \delta_{n-1,j}.$$

For $j < n-1$, we have $b_{n,j} = 0$ and $\delta_{n-1,j} = 0$, so $b_{n-1,j} = 0$. For $j = n-1$, we have $b_{n,n-1} = 0$ and $\delta_{n-1,n-1} = 1$, so $b_{n-1,n-1} = a_{n-1,n-1}^{-1} = \lambda_{n-1}^{-1}$. Assume inductively that for some $k \in \{1, \ldots, n-1\}$, we have $b_{ij} = 0$ for all $i > j$ and $b_{ii} = \lambda_i^{-1}$ for all $i \geq k+1$. Now, consider the k-th row. We have

$$a_{kk} b_{kj} + a_{k,k+1} b_{k+1,j} + \cdots + a_{kn} b_{nj} = \delta_{kj}.$$

For $j < k$, we have $b_{ij} = 0$ for all $i > j \geq k$, so the left-hand side reduces to $a_{kk} b_{kj} = 0$, hence $b_{kj} = 0$. For $j = k$, we have $b_{ij} = 0$ for all $i > k$, so the left-hand side reduces to $a_{kk} b_{kk} = 1$, hence $b_{kk} = a_{kk}^{-1} = \lambda_k^{-1}$. By induction, we conclude that B is upper-triangular with $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ on the diagonal. □

Solution 5C.4 Consider the linear operator $T \in \text{End}(\mathbb{R}^2)$ with matrix representation

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

with respect to the standard basis. A is invertible with inverse

$$A^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

□

Solution 5C.5 Consider the linear operator $T \in \text{End}(\mathbb{R}^2)$ with matrix representation

$$A = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$

with respect to the standard basis. Since $T(1,-1) = (0,0)$, T is not invertible. □

Solution 5C.6 Recall that every operator on a finite-dimensional complex vector space has an upper-triangular matrix with respect to some basis. Let $v_1, \ldots, v_n$ be such a basis. Then, for each $k \in \{1, \ldots, n\}$,

$$Tv_j \in \text{span}(v_1, \ldots, v_j) \subseteq \text{span}(v_1, \ldots, v_k)$$

for all $j \leq k$. Hence, $\text{span}(v_1, \ldots, v_k)$ is a k-dimensional subspace that is invariant under T. □

Solution 5C.7

(a) If $v = 0$, then $p(T)v = 0$ for any polynomial p, so we can take $p_v(x) = 1$. Otherwise, suppose $v \neq 0$. Consider the list $v, Tv, T^2v, \ldots, T^{\dim V}v$. Since V is finite-dimensional, this list is linearly dependent. Then there exist scalars $a_0, \ldots, a_m$ not all zero such that

$$a_0 v + a_1 Tv + \cdots + a_m T^m v = 0.$$

Here, m denotes the smallest integer such that there exist scalars $a_0, \ldots, a_m$ not all zero satisfying the above equation. Since m is minimal, we have $a_m \neq 0$. Let $p_v(x)$ be the monic polynomial after dividing by a_m:

$$p_v(x) = x^m + \frac{a_{m-1}}{a_m} x^{m-1} + \cdots + \frac{a_0}{a_m}.$$

Then $p_v(T)v = 0$. Suppose q is another monic polynomial of smallest degree such that $q(T)v = 0$. Then $p_v - q$ is a polynomial of smaller degree than p_v such that $(p_v - q)(T)v = 0$, which contradicts the minimality of p_v.

(b) Let m_T be the minimal polynomial of T. Then $m_T(T) = 0$, so $m_T(T)v = 0$. By the division algorithm for polynomials, there exist unique polynomials q_0 and r such that

$$m_T = p_v q_0 + r,$$

where $r = 0$ or $\deg r < \deg p_v$. Then

$$0 = m_T(T)v = (p_v q_0 + r)(T)v = p_v(T)q_0(T)v + r(T)v = q_0(T)p_v(T)v + r(T)v = r(T)v,$$

since $p_v(T)v = 0$. If $r \neq 0$, then r is a polynomial of smaller degree than p_v such that $r(T)v = 0$, which contradicts the minimality of p_v. Therefore, $r = 0$, and we have $m_T = p_v q_0$, which shows that m_T is a polynomial multiple of p_v.

□

Solution 5C.8

(a) Suppose there exists a basis of V with respect to which T has an upper-triangular matrix. Then the diagonal entries of this matrix are the eigenvalues of T. Let $p(z) = z^2 + 2z + 2$. Then $p(T)$ also has an upper-triangular matrix with respect to the same basis, and the diagonal entries of this matrix are $p(\lambda)$, where λ is an eigenvalue of T. Since $p(T)v = 0$ for some nonzero vector v, $p(T)$ is not injective. Then $p(T)$ is not invertible, so $p(\lambda) = 0$ for some eigenvalue $\lambda \in \mathbb{R}$ of T. However, $p(z)$ has no real roots, contradicting that $p(\lambda) = 0$ for some real eigenvalue λ.

(b) Utilizing the same deductions as in part (a), we conclude that there exists an eigenvalue λ of T such that $p(\lambda) = 0$. The roots of $p(z)$ are $-1 + i$ and $-1 - i$, so either $-1 + i$ or $-1 - i$ must be an eigenvalue of T. Since the diagonal entries of an upper-triangular matrix are the eigenvalues of the corresponding operator, either $-1 + i$ or $-1 - i$ must appear on the diagonal of A.

□

Solution 5C.9 Since B represents a linear operator on a finite-dimensional complex vector space, let T be the linear operator corresponding to B, i.e., $B = \text{Mat}(T, (e_1, \ldots, e_n))$, where $(e_1, \ldots, e_n)$ is the standard basis of $\mathbb{C}^n$. Since T has an upper-triangular matrix with respect to some basis of the vector space, let A be the change of basis matrix from the new basis to the standard basis. Then $A^{-1}BA$ is the matrix representation of T with respect to this new basis, which is upper-triangular. □

Solution 5C.10

- Suppose the matrix of T with respect to $v_1, \ldots, v_n$ is lower triangular. Fix some $k \in \{1, \ldots, n\}$. For $j \in \{1, \ldots, n\}$,

$$Tv_j \in \text{span}(v_j, \ldots, v_n)$$

  because the matrix is lower triangular. If $j \geq k$, then $\text{span}(v_j, \ldots, v_n) \subseteq \text{span}(v_k, \ldots, v_n)$. Therefore, $Tv_j \in \text{span}(v_k, \ldots, v_n)$ for all $j \geq k$, which shows that $\text{span}(v_k, \ldots, v_n)$ is invariant under T.

- Suppose $\text{span}(v_k, \ldots, v_n)$ is invariant under T for each $k = 1, \ldots, n$. Then for each k, we have $Tv_k \in \text{span}(v_k, \ldots, v_n)$, which shows that Tv_k can be expressed as a linear combination of $v_k, \ldots, v_n$.

- Suppose $Tv_k \in \text{span}(v_k, \ldots, v_n)$ for each $k = 1, \ldots, n$. Then

$$Tv_k = \sum_{j=k}^{n} A_{j,k} v_j$$

  for some scalars $A_{j,k}$. Then $A_{i,k} = 0$ for all $i < k$, which shows that the matrix of T with respect to $v_1, \ldots, v_n$ is lower triangular.

□

Solution 5C.11 Since T has an upper-triangular matrix with respect to some basis of V, let $v_1, \ldots, v_n$ be such a basis. Then $\operatorname{span}(v_1, \ldots, v_k)$ is invariant under T for each $k = 1, \ldots, n$. Consider the reverse basis $w_1 = v_n, w_2 = v_{n-1}, \ldots, w_n = v_1$. Then $\operatorname{span}(w_k, \ldots, w_n) = \operatorname{span}(v_1, \ldots, v_{n-k+1})$ is invariant under T for each $k = 1, \ldots, n$. By Exercise 5C10, the matrix of T with respect to the basis $w_1, \ldots, w_n$ is lower triangular. □

Solution 5C.12

Axler Note. *The quotient operator T/U was defined in Exercise 5A38.*

(a) Since T has an upper-triangular matrix with respect to some basis of V, then its minimal polynomial is a product of linear factors. Since U is invariant under T, the minimal polynomial of $T|_U$ divides the minimal polynomial of T, and thus is also a product of linear factors. Therefore, $T|_U$ has an upper-triangular matrix with respect to some basis of U.

(b) Since the minimal polynomial of T is a polynomial multiple of the minimal polynomial of T/U, the same argument as in part (a) shows that T/U has an upper-triangular matrix with respect to some basis of V/U.

□

Solution 5C.13 Let $u_1, \ldots, u_m$ be a basis of U with respect to which $T|_U$ has an upper-triangular matrix. Let $v_1 + U, \ldots, v_n + U$ be a basis of V/U with respect to which T/U has an upper-triangular matrix. Then the list $u_1, \ldots, u_m, v_1, \ldots, v_n$ is a basis of V. Since $T|_U$ has an upper-triangular matrix with respect to the basis $u_1, \ldots, u_m$, we have

$$Tu_j = \sum_{k=1}^{j} A_{k,j} u_k$$

for some scalars $A_{k,j}$ and for each $j = 1, \ldots, m$. Since T/U has an upper-triangular matrix with respect to the basis $v_1 + U, \ldots, v_n + U$, we have

$$(T/U)(v_j + U) = \sum_{k=1}^{j} B_{k,j}(v_k + U)$$

for some scalars $B_{k,j}$ and for each $j = 1, \ldots, n$. This implies that

$$Tv_j - \sum_{k=1}^{j} B_{k,j} v_k \in U,$$

so there exist scalars $C_{l,j}$ such that

$$Tv_j - \sum_{k=1}^{j} B_{k,j} v_k = \sum_{l=1}^{m} C_{l,j} u_l.$$

Therefore,

$$Tv_j = \sum_{l=1}^{m} C_{l,j} u_l + \sum_{k=1}^{j} B_{k,j} v_k,$$

which shows that the matrix of T with respect to the basis $u_1, \dots, u_m, v_1, \dots, v_n$ is upper-triangular. □

Solution 5C.14

Manual Remark. *This can be argued by noting that* $\mathrm{Mat}(T^t) = \mathrm{Mat}(T)^t$, *so the matrix of* T' *with respect to the dual basis is lower-triangular. Then the reversed dual basis gives an upper-triangular matrix for* T'.

By Exercise 5B28, the minimal polynomials of T and T' are the same. Therefore, the following are equivalent:

- T has an upper-triangular matrix with respect to some basis of V.
- The minimal polynomial of T is a product of linear factors.
- The minimal polynomial of T' is a product of linear factors.
- T' has an upper-triangular matrix with respect to some basis of V'. □

§5D

Solution 5D.1

Manual Remark. *The roots* ω *in part (b) are the roots obtained in Exercise 1C6.*

(a) If $T^4 = I$, then the minimal polynomial of T divides $z^4 - 1 = (z-1)(z+1)(z-i)(z+i)$, which has no repeated roots. Hence, the minimal polynomial of T has no repeated roots, so T is diagonalizable.

(b) If $T^4 = T$, then the minimal polynomial of T divides $z^4 - z = z(z-1)(z-\omega)(z-\omega^2)$. This polynomial has no repeated roots, so the minimal polynomial of T has no repeated roots, and hence T is diagonalizable.

(c) Let $T \in \mathrm{End}(\mathbb{C}^2)$ be defined by

$$T(z_1, z_2) = (z_2, 0).$$

Then $T^2(z_1, z_2) = (0, 0)$, so $T^4 = T^2$. However, T is not diagonalizable because the only eigenvalue of T is 0, and $\dim E(0, T) = 1 < 2 = \dim \mathbb{C}^2$.

□

Solution 5D.2 Suppose A is a diagonal matrix of T with respect to some basis of V. Let $\lambda \in \mathbb{F}$. Let m denote the number of times λ appears on the diagonal of A. Then $A - \lambda I$ is a diagonal matrix of $T - \lambda I$ with respect to the same basis, with exactly m zeros on the diagonal. A diagonal matrix with m zeros on the diagonal has null space of dimension m, so $\dim \ker(T - \lambda I) = m$. Since $E(\lambda, T) = \ker(T - \lambda I)$, we have $\dim E(\lambda, T) = m$. □

Solution 5D.3 Suppose T is diagonalizable. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T. Then

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T).$$

If 0 is not an eigenvalue of T, then $\ker T = \{0\}$ and $\operatorname{im} T = V$, so $V = \ker T \oplus \operatorname{im} T$. If 0 is an eigenvalue of T, then we may assume $\lambda_1 = 0$. Let

$$U = E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T).$$

Since $\ker T = E(0, T)$, we have $V = \ker T \oplus U$. It remains to show that $U = \operatorname{im} T$. Let $u \in U$. Then $u = u_2 + \cdots + u_m$ for some $u_k \in E(\lambda_k, T)$. Since $\lambda_k \neq 0$ for $k = 2, \ldots, m$, we have $u_k = \lambda_k^{-1} T(u_k)$ for each $k = 2, \ldots, m$. Hence,

$$u = u_2 + \cdots + u_m = \lambda_2^{-1} T(u_2) + \cdots + \lambda_m^{-1} T(u_m) = T(\lambda_2^{-1} u_2 + \cdots + \lambda_m^{-1} u_m) \in \operatorname{im} T.$$

Thus, $U \subseteq \operatorname{im} T$.

Now let $v \in \operatorname{im} T$. Note that eigenspaces are invariant as follows: If $u \in E(\lambda_k, T)$, then $T(u) = \lambda_k u \in E(\lambda_k, T)$. Since U is a direct sum of eigenspaces, it is invariant under T. Then $v = T(w)$ for some $w \in V$. Since $V = \ker T \oplus U$, we may write $w = w_1 + w_2$ for some $w_1 \in \ker T$ and $w_2 \in U$. Then

$$v = T(w) = T(w_1 + w_2) = T(w_1) + T(w_2) = 0 + T(w_2) = T(w_2) \in U.$$

Thus, $\operatorname{im} T \subseteq U$. Since $U \subseteq \operatorname{im} T$ and $\operatorname{im} T \subseteq U$, we have $U = \operatorname{im} T$. Therefore, $V = \ker T \oplus U = \ker T \oplus \operatorname{im} T$. □

Solution 5D.4
(a) $\Longrightarrow$ (b): If $V = \ker T \oplus \operatorname{im} T$, then $V = \ker T + \operatorname{im} T$.
(b) $\Longrightarrow$ (c): If $V = \ker T + \operatorname{im} T$, then $\dim V = \dim \ker T + \dim \operatorname{im} T - \dim(\ker T \cap \operatorname{im} T)$. Since $\dim V = \dim \ker T + \dim \operatorname{im} T$ by the fundamental theorem of linear maps, we have $\dim(\ker T \cap \operatorname{im} T) = 0$, so $\ker T \cap \operatorname{im} T = \{0\}$.
(c) $\Longrightarrow$ (a): If $\ker T \cap \operatorname{im} T = \{0\}$, then

$$\begin{aligned} \dim(\ker T + \operatorname{im} T) &= \dim \ker T + \dim \operatorname{im} T - \dim(\ker T \cap \operatorname{im} T) \\ &= \dim \ker T + \dim \operatorname{im} T. \end{aligned}$$

By the fundamental theorem of linear maps, we have $\dim V = \dim \ker T + \dim \operatorname{im} T$. Hence,

$$\dim(\ker T + \operatorname{im} T) = \dim V.$$

Since $\ker T + \operatorname{im} T \subseteq V$ and $\dim(\ker T + \operatorname{im} T) = \dim V$, we have $\ker T + \operatorname{im} T = V$. Along with the assumption that $\ker T \cap \operatorname{im} T = \{0\}$, we have $V = \ker T \oplus \operatorname{im} T$. □

Solution 5D.5 Suppose T is diagonalizable. Let $\lambda \in \mathbb{C}$. Since T is diagonalizable, we have

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$$

for some distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ of T. If λ is not an eigenvalue of T, then $\ker(T - \lambda I) = \{0\}$ and $\operatorname{im}(T - \lambda I) = V$, so

$$V = \ker(T - \lambda I) \oplus \operatorname{im}(T - \lambda I).$$

If λ is an eigenvalue of T, then we may assume $\lambda_1 = \lambda$. Let

$$U = E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T).$$

Then $V = E(\lambda, T) \oplus U$. It remains to show that $U = \operatorname{im}(T - \lambda I)$.
By dimension counting, we have

$$\dim U = \dim V - \dim E(\lambda, T) = \dim V - \dim \ker(T - \lambda I) = \dim \operatorname{im}(T - \lambda I).$$

Since each $u_k \in E(\lambda_k, T)$ satisfies $T(u_k) = \lambda_k u_k$, we have

$$(T - \lambda I)(u_k) = (\lambda_k - \lambda)u_k \in E(\lambda_k, T).$$

To see that $T - \lambda I$ is injective on U, suppose $u \in U$ and $(T - \lambda I)(u) = 0$. Then $u = u_2 + \cdots + u_m$ for some $u_k \in E(\lambda_k, T)$, and

$$\begin{aligned} 0 &= (T - \lambda I)(u) \\ &= (T - \lambda I)(u_2 + \cdots + u_m) \\ &= (T - \lambda I)(u_2) + \cdots + (T - \lambda I)(u_m) \\ &= (\lambda_2 - \lambda)u_2 + \cdots + (\lambda_m - \lambda)u_m. \end{aligned}$$

Since $\lambda_2, \ldots, \lambda_m$ are distinct from λ, we have $\lambda_k - \lambda \neq 0$ for each $k = 2, \ldots, m$. Since the sum above is a direct sum, we have $(\lambda_k - \lambda)u_k = 0$ for each $k = 2, \ldots, m$, so $u_k = 0$ for each $k = 2, \ldots, m$. Hence, $u = u_2 + \cdots + u_m = 0$, so $T - \lambda I$ is injective on U. Since $\dim U = \dim \operatorname{im}(T - \lambda I)$ and U is invariant under $T - \lambda I$, we have $U = \operatorname{im}(T - \lambda I)$.

Conversely, suppose $V = \ker(T - \lambda I) \oplus \operatorname{im}(T - \lambda I)$ for every $\lambda \in \mathbb{C}$. Then $\ker(T - \lambda I) \cap \operatorname{im}(T - \lambda I) = \{0\}$ for every $\lambda \in \mathbb{C}$. Note that $T - \lambda I$ restricted to $\operatorname{im}(T - \lambda I)$ is injective, so λ is not an eigenvalue of $T|_{\operatorname{im}(T-\lambda I)}$. Then the minimal polynomial of $T|_{\operatorname{im}(T-\lambda I)}$ does not have λ as a root. Suppose $(z - \lambda)^2$ divides the minimal polynomial of T. Then there is some $v \in V$ such that $(T - \lambda I)^2 v = 0$, but $(T - \lambda I)v \neq 0$. Then $(T - \lambda I)v \in \operatorname{im}(T - \lambda I)$, and $v \in \ker(T - \lambda I) \cap \operatorname{im}(T - \lambda I) = \{0\}$, which is a contradiction. Hence, $(z - \lambda)^2$ does not divide the minimal polynomial of T for any $\lambda \in \mathbb{C}$, so the minimal polynomial of T has no repeated roots. Since the minimal polynomial of T has no repeated roots, T is diagonalizable. □

Solution 5D.6 Suppose $T - 2I$ and $T - 6I$ are not invertible. Then 2 and 6 are eigenvalues of T, so $\dim E(2, T) \geq 1$ and $\dim E(6, T) \geq 1$. Since 8 is also an eigenvalue of T, we have

$$\dim E(2, T) + \dim E(6, T) + \dim E(8, T) \geq 1 + 1 + 4 = 6 > 5 = \dim \mathbb{F}^5,$$

which is a contradiction. Hence, $T - 2I$ or $T - 6I$ is invertible. □

Solution 5D.7 Suppose $v \in E(\lambda, T)$. Then $T(v) = \lambda v$. Since T is invertible, we have $v = T^{-1}(\lambda v) = \lambda T^{-1}(v)$, so $T^{-1}(v) = \frac{1}{\lambda} v$. Hence,

$$v \in E(\lambda^{-1}, T^{-1}),$$

so $E(\lambda, T) \subseteq E(\lambda^{-1}, T^{-1})$.
Now suppose $v \in E(\lambda^{-1}, T^{-1})$. Then

$$T^{-1}(v) = \lambda^{-1}v \implies v = T(\lambda^{-1}v) \implies v = \lambda^{-1}T(v) \implies T(v) = \lambda v.$$

Hence, $v \in E(\lambda, T)$, so $E(\lambda^{-1}, T^{-1}) \subseteq E(\lambda, T)$. Then $E(\lambda, T) = E(\lambda^{-1}, T^{-1})$. □

Solution 5D.8 For every nonzero eigenvalue λ of T and every $v \in E(\lambda, T)$, we have $v = T(\lambda^{-1}v) \in \operatorname{im} T$, so $E(\lambda, T) \subseteq \operatorname{im} T$. Since eigenspaces corresponding to distinct eigenvalues are linearly independent, we have

$$E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T) \subseteq \operatorname{im} T.$$

It follows that

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) = \dim(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)) \leq \dim \operatorname{im} T. \quad \square$$

Solution 5D.9 Since R and T each have 3 distinct eigenvalues, they are both diagonalizable. Let r_2, r_6, r_7 be eigenvectors of R with $Rr_k = kr_k$ for $k = 2, 6, 7$, and let t_2, t_6, t_7 be eigenvectors of T with $Tt_k = kt_k$ for $k = 2, 6, 7$. Each list of 3 eigenvectors is linearly independent, so they each form a basis of $\mathbb{F}^3$. Define $S \in \operatorname{End}(\mathbb{F}^3)$ by $S(r_k) = t_k$ for $k = 2, 6, 7$. Then S is invertible because it maps a basis of $\mathbb{F}^3$ to a basis of $\mathbb{F}^3$. For each $k = 2, 6, 7$, we have

$$S^{-1}TS(r_k) = S^{-1}T(t_k) = S^{-1}(kt_k) = kS^{-1}(t_k) = kr_k = R(r_k).$$

Since $S^{-1}TS$ and R agree on a basis of $\mathbb{F}^3$, we have $R = S^{-1}TS$. □

Solution 5D.10 Define R to be the operator whose matrix with respect to the standard basis of $\mathbb{F}^4$ has diagonal entries $2, 6, 7, 7$ and all other entries 0. Then R has eigenvalues $2, 6, 7$ and no other eigenvalues. Define T to be the operator whose matrix with respect to the standard basis of $\mathbb{F}^4$ has diagonal entries $2, 6, 7, 2$ and all other entries 0. Then T has eigenvalues $2, 6, 7$ and no other eigenvalues. If there did exist an invertible operator $S \in \operatorname{End}(\mathbb{F}^4)$ such that $R = S^{-1}TS$, then

$$Rv = S^{-1}TSv = \lambda v \implies T(Sv) = \lambda Sv$$

for every eigenvalue λ of R and every eigenvector $v \in E(\lambda, R)$. Hence, the map $S|_{E(\lambda,R)}$ maps $E(\lambda, R)$ to $E(\lambda, T)$. Since S is invertible, $S|_{E(\lambda,R)}$ is injective, so $\dim E(\lambda, R) \leq \dim E(\lambda, T)$. However, $\dim E(7, R) = 2 > 1 = \dim E(7, T)$, so there does not exist an invertible operator $S \in \operatorname{End}(\mathbb{F}^4)$ such that $R = S^{-1}TS$. □

Solution 5D.11 Consider the operator $T \in \operatorname{End}(\mathbb{C}^3)$ whose matrix with respect to the standard basis of $\mathbb{C}^3$ is

$$\begin{pmatrix} 6 & 1 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

Observe that 6 and 7 are the only eigenvalues of T since the matrix is upper triangular with diagonal entries 6, 6, 7. If T had a diagonal matrix with respect to some basis of $\mathbb{C}^3$, then $\dim E(6,T) + \dim E(7,T) = 3$. However, $E(6,T)$ is the null space of $T-6I$, which has matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that the null space of this matrix is spanned by $(1,0,0)$, so $\dim E(6,T) = 1$. Similarly, $E(7,T)$ is the null space of $T-7I$, which has matrix

$$\begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Observe that the null space of this matrix is spanned by $(0,0,1)$, so $\dim E(7,T) = 1$. Hence, $\dim E(6,T) + \dim E(7,T) = 1 + 1 = 2 < 3$, so T does not have a diagonal matrix with respect to any basis of $\mathbb{C}^3$. □

Solution 5D.12 Since T cannot be diagonalized, it cannot have 3 distinct eigenvalues. Then 6 and 7 must be the only eigenvalues of T. In particular, 8 is not an eigenvalue of T, so $T-8I$ is surjective. Then there exists $(z_1,z_2,z_3) \in \mathbb{C}^3$ such that

$$(T-8I)(z_1,z_2,z_3) = (6,7,13) \implies T(z_1,z_2,z_3) = (6+8z_1, 7+8z_2, 13+8z_3). \quad \square$$

Solution 5D.13 Let A be a diagonal matrix with distinct entries on the diagonal, and let B be a matrix of the same size as A. Suppose $AB = BA$. Let $A_{j,j}$ denote the j-th diagonal entry of A, and let $B_{j,k}$ denote the entry of B in row j and column k. Then

$$(AB)_{j,k} = A_{j,j}B_{j,k}, \quad (BA)_{j,k} = B_{j,k}A_{k,k}.$$

Since $AB = BA$, we have

$$A_{j,j}B_{j,k} = B_{j,k}A_{k,k}.$$

If $j \neq k$, then $A_{j,j} - A_{k,k} \neq 0$ because the diagonal entries of A are distinct. Hence,

$$(A_{j,j} - A_{k,k})B_{j,k} = 0 \implies B_{j,k} = 0.$$

Thus, all off-diagonal entries of B are zero, so B is a diagonal matrix.
Conversely, if B is a diagonal matrix, then it is straightforward to verify that $AB = BA$. □

Solution 5D.14

(a) Consider the operator $T \in \text{End}(\mathbb{C}^2)$ defined by

$$T(z_1,z_2) = (z_2,0).$$

Then $T^2(z_1,z_2) = (0,0)$, so T^2 is diagonalizable. However, T is not diagonalizable because the only eigenvalue of T is 0, and $\dim E(0,T) = 1 < 2 = \dim \mathbb{C}^2$.

(b) Suppose T is diagonalizable. Then a basis of V consists of eigenvectors of T, and each eigenvector of T is also an eigenvector of T^k. Hence, T^k is diagonalizable.

Conversely, suppose T^k is diagonalizable. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T^k. Then the minimal polynomial of T^k has no repeated roots, so we may write

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m).$$

Consider the polynomial

$$q(z) = (z^k - \lambda_1) \cdots (z^k - \lambda_m).$$

Note that $q(T) = p(T^k) = 0$, so the minimal polynomial of T divides q. Since T is invertible, T^k is also invertible. Then $\lambda_j \neq 0$ for each $j = 1, \ldots, m$, so q has no repeated roots. Hence, the minimal polynomial of T has no repeated roots, so T is diagonalizable.

□

Solution 5D.15

Axler Note. *The* greatest common divisor *of p and p' is the monic polynomial q of largest degree such that p and p' are both polynomial multiples of q. The Euclidean algorithm for polynomials (look it up) can quickly determine the greatest common divisor of two polynomials, without requiring any information about the zeros of the polynomials. Thus, the equivalence of (a) and (d) above shows that we can determine whether T is diagonalizable without knowing anything about the zeros of p.*

(a) $\implies$ (b) Suppose T is diagonalizable. Then the minimal polynomial of T has no repeated roots, so there does not exist $\lambda \in \mathbb{C}$ such that p is a polynomial multiple of $(z - \lambda)^2$.
(b) $\implies$ (c) Suppose there does not exist $\lambda \in \mathbb{C}$ such that p is a polynomial multiple of $(z - \lambda)^2$. Then p has no repeated roots, so we may let

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m)$$

for some distinct $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$. Then evaluating the derivative p' at λ_k gives

$$p'(\lambda_k) = \prod_{j \neq k} (\lambda_k - \lambda_j) \neq 0$$

by the distinctness of $\lambda_1, \ldots, \lambda_m$. Hence, p and p' have no zeros in common.
(c) $\implies$ (d) Suppose p and its derivative p' have no zeros in common. Then the greatest common divisor of p and p' has no zeros, so it is a constant polynomial.
(d) $\implies$ (a) Suppose the greatest common divisor of p and p' is the constant polynomial 1. If p had a repeated root $\lambda \in \mathbb{C}$, then p would be a polynomial multiple of $(z - \lambda)^2$, so p' would be a polynomial multiple of $(z - \lambda)$. Hence, p and p' would have a zero in common, which is a contradiction. Thus, p has no repeated roots, so T is diagonalizable. □

Solution 5D.16 Suppose U is invariant under T. Let $U_k = U \cap E(\lambda_k, T)$ for each $k = 1, \ldots, m$. Then $U_k \subseteq E(\lambda_k, T)$ for each k. For each k, define the polynomial p_k by

$$p_k(z) = \prod_{j \neq k} \frac{z - \lambda_j}{\lambda_k - \lambda_j}.$$

Then $p_k(\lambda_j) = 0$ for $j \neq k$ and $p_k(\lambda_k) = 1$. Let $u \in U$ with

$$u = u_1 + \cdots + u_m, \quad u_k \in E(\lambda_k, T).$$

Then $p_k(T)(u) = u_k \in U$ for each k, so $u_k \in U_k$ for each k. Hence, $U = U_1 \oplus \cdots \oplus U_m$. Conversely, suppose there exist subspaces $U_1, \ldots, U_m$ of V such that $U_k \subseteq E(\lambda_k, T)$ for each k and $U = U_1 \oplus \cdots \oplus U_m$. Then for each $u \in U$, we have

$$u = u_1 + \cdots + u_m, \quad u_k \in U_k.$$

Then

$$T(u) = T(u_1 + \cdots + u_m) = T(u_1) + \cdots + T(u_m) = \lambda_1 u_1 + \cdots + \lambda_m u_m \in U,$$

so U is invariant under T. □

Solution 5D.17 Let $n = \dim V$. Let $\{v_1, \ldots, v_n\}$ be a basis of V. For each $j, k \in \{1, \ldots, n\}$, define $T_{j,k} \in \mathrm{End}(V)$ by

$$T_{j,k}(v_\ell) = \begin{cases} v_j & \text{if } \ell = k, \\ 0 & \text{if } \ell \neq k. \end{cases}$$

If $j = k$, let $D_j = T_{j,j}$. Then D_j is diagonalizable because $D_j(v_j) = v_j$ and $D_j(v_\ell) = 0$ for $\ell \neq j$. If $j \neq k$, let $N_{j,k} = D_j + T_{j,k}$. Then $N_{j,k}$ is diagonalizable because $N_{j,k}(v_j) = v_j$, $N_{j,k}(v_k) = v_j$, and $N_{j,k}(v_\ell) = 0$ for $\ell \neq j, k$. Then the eigenvalues of $N_{j,k}$ are 0 with corresponding eigenvector $v_j - v_k$, and 1 with $n - 2$ corresponding eigenvectors v_ℓ for $\ell \neq j, k$, so $N_{j,k}$ is diagonalizable.
Note that the set

$$\{D_1, \ldots, D_n\} \cup \{N_{j,k} : j, k \in \{1, \ldots, n\}, j \neq k\}$$

spans $\mathrm{End}(V)$. For any $T \in \mathrm{End}(V)$, we have

$$T = \sum_{j,k=1}^{n} \alpha_{j,k} T_{j,k} = \sum_{j \neq k} \alpha_{j,k}(N_{j,k} - D_j) + \sum_{j=1}^{n} \alpha_{j,j} D_j,$$

which shows that the set above spans $\mathrm{End}(V)$.

Moreover, note that $\dim \mathrm{End}(V) = n^2$ and the set above has $n + n(n-1) = n^2$ elements, so it is a basis of $\mathrm{End}(V)$ consisting of diagonalizable operators. □

Solution 5D.18 Since T is diagonalizable, V has a basis $v_1, \ldots, v_n$ of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues of T. Note that $v_1 + U, \ldots, v_n + U$ span V/U. Moreover, for each j, we have

$$(T/U)(v_j + U) = T(v_j) + U = \lambda_j v_j + U = \lambda_j (v_j + U),$$

so V/U is spanned by eigenvectors of T/U. We may reduce this spanning set to a basis of V/U consisting of eigenvectors of T/U, so T/U is diagonalizable. □

Solution 5D.19

Axler Note. *See Exercise 5C13 for an analogous statement about upper-triangular matrices.*

This statement is false. Consider the operator $T \in \text{End}(\mathbb{F}^2)$ whose matrix with respect to the standard basis of $\mathbb{F}^2$ is

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Let $U = \text{span}\{(1, 0)\}$. Then U is invariant under T since $T(1, 0) = (0, 0) \in U$. The operator $T|_U$ is diagonalizable because it is the zero operator on a one-dimensional space. The quotient space V/U is one-dimensional, so T/U is diagonalizable. However, T is not diagonalizable because the only eigenvalue of T is 0, and $\dim E(0, T) = 1 < 2 = \dim \mathbb{F}^2$. □

Solution 5D.20 Suppose T is diagonalizable. Then there exists a basis of V consisting of eigenvectors of T. Let $\{v_1, \ldots, v_n\}$ be a basis of V consisting of eigenvectors of T, and let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Let $\{\phi_1, \ldots, \phi_n\}$ be the dual basis of V' corresponding to $\{v_1, \ldots, v_n\}$. Then for each $k = 1, \ldots, n$, we have

$$T'(\phi_k)(v_j) = \phi_k(T(v_j)) = \phi_k(\lambda_j v_j) = \lambda_j \phi_k(v_j) = \begin{cases} \lambda_k & j = k, \\ 0 & j \neq k, \end{cases}$$

so $T'(\phi_k) = \lambda_k \phi_k$. Hence, $\{\phi_1, \ldots, \phi_n\}$ is a basis of V' consisting of eigenvectors of T', so T' is diagonalizable.

Conversely, suppose T' is diagonalizable. Then there exists a basis of V' consisting of eigenvectors of T'. Using the same reasoning as above, $(T')'$ is diagonalizable with the same eigenvalues as T'. From Exercise 3F32, we know that V is isomorphic to V'' with the map $\Lambda : V \to V''$ defined by

$$\Lambda(v)(\phi) = \phi(v)$$

for each $v \in V$ and $\phi \in V'$. Then $(T')'(\Lambda(v)) = \Lambda(T(v))$ for each $v \in V$, so T has a basis of eigenvectors corresponding to the same eigenvalues as T', so T is diagonalizable. □

Solution 5D.21

Axler Note. *Each F_n is a nonnegative integer, even though the right side of the formula in (d) does not look like an integer. The number*

$$\frac{1+\sqrt{5}}{2}$$

is called the ***golden ratio****.*

Manual Remark. *Part (e) uses the triangle inequality and creating "zeros" in the right places to show that the second term in the formula for F_n is less than $1/2$ in absolute value.*

(a) We prove the statement by induction on n. For the base case $n = 0$, we have $T^0(0,1) = (0,1) = (F_0, F_1)$. Now suppose the statement holds for some nonnegative integer n. Then

$$\begin{aligned} T^{n+1}(0,1) &= T(T^n(0,1)) \\ &= T(F_n, F_{n+1}) \\ &= (F_{n+1}, F_n + F_{n+1}) \\ &= (F_{n+1}, F_{n+2}), \end{aligned}$$

so the statement holds for $n+1$.

(b) Suppose $T(x,y) = \lambda(x,y)$ for some $\lambda \in \mathbb{R}$ and $(x,y) \in \mathbb{R}^2$ with $(x,y) \neq (0,0)$. Then

$$(y, x+y) = (\lambda x, \lambda y),$$

so $y = \lambda x$ and $x + y = \lambda y$. We then have

$$x + \lambda x = \lambda^2 x \implies (1 + \lambda - \lambda^2)x = 0.$$

If $x = 0$, then $y = \lambda x = 0$, which is a contradiction. Hence, $x \neq 0$, so we have

$$\lambda^2 - \lambda - 1 = 0 \implies \lambda = \frac{1 \pm \sqrt{5}}{2}.$$

(c) Let

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \quad \lambda_2 = \frac{1-\sqrt{5}}{2}.$$

Using $y = \lambda x$, we have the basis $\{(1,\lambda_1), (1,\lambda_2)\}$ of $\mathbb{R}^2$ consisting of eigenvectors of T.

(d) Let v_1 and v_2 be the eigenvectors of T corresponding to the eigenvalues λ_1 and λ_2, respectively. Since v_1, v_2 form a basis of $\mathbb{R}^2$, we may write $(0,1) = \alpha v_1 + \beta v_2$ for some $\alpha, \beta \in \mathbb{R}$:

$$(0,1) = \alpha\left(1, \frac{1+\sqrt{5}}{2}\right) + \beta\left(1, \frac{1-\sqrt{5}}{2}\right) = (\alpha+\beta, \alpha\lambda_1 + \beta\lambda_2).$$

Then $\alpha + \beta = 0$ and $\alpha\lambda_1 + \beta\lambda_2 = 1$, so $\beta = -\alpha$ and $\alpha(\lambda_1 - \lambda_2) = 1$. Note that $\lambda_1 - \lambda_2 = \sqrt{5}$, so $\alpha = 1/\sqrt{5}$ and $\beta = -1/\sqrt{5}$. Then

$$\begin{aligned} T^n(0,1) &= T^n\left(\frac{1}{\sqrt{5}}v_1 - \frac{1}{\sqrt{5}}v_2\right) \\ &= \frac{1}{\sqrt{5}}T^n(v_1) - \frac{1}{\sqrt{5}}T^n(v_2) \\ &= \frac{1}{\sqrt{5}}\lambda_1^n v_1 - \frac{1}{\sqrt{5}}\lambda_2^n v_2 \\ &= \frac{1}{\sqrt{5}}\left(\lambda_1^n - \lambda_2^n, \lambda_1^{n+1} - \lambda_2^{n+1}\right). \end{aligned}$$

Hence, we have

$$F_n = \frac{1}{\sqrt{5}}\left(\lambda_1^n - \lambda_2^n\right) = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

(e) Observe that $\frac{1}{\sqrt{5}} < \frac{1}{2}$. Moreover,

$$|\lambda_2| = \frac{\sqrt{5}-1}{2} < 1$$

since $\sqrt{5} < 3$. Then for all $n \geq 0$, we have $|\lambda_2^n| \leq 1$, so

$$\left|\frac{1}{\sqrt{5}}\lambda_2^n\right| < \frac{1}{2} \implies \left|\frac{1}{\sqrt{5}}\lambda_1^n - F_n\right| < \frac{1}{2}.$$

Let m be an integer such that $m \neq F_n$. Then $|m - F_n| \geq 1$, so

$$\left|\frac{1}{\sqrt{5}}\lambda_1^n - m\right| = \left|\frac{1}{\sqrt{5}}\lambda_1^n - F_n + F_n - m\right| \geq |F_n - m| - \left|\frac{1}{\sqrt{5}}\lambda_1^n - F_n\right| > 1 - \frac{1}{2} = \frac{1}{2}.$$

Since the distance from $\frac{1}{\sqrt{5}}\lambda_1^n$ to any integer other than F_n is greater than $1/2$, we conclude that F_n is the integer closest to $\frac{1}{\sqrt{5}}\lambda_1^n$.

□

Solution 5D.22 By Theorem 5.67, each eigenvalue λ of T lies in at least one of the Gershgorin disks

$$D_j = \left\{z \in \mathbb{C} : |z - A_{j,j}| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|\right\}.$$

If $\lambda = 0$ is an eigenvalue of T, then 0 lies in at least one of the Gershgorin disks D_j. However, by the given inequality, we have

$$|0 - A_{j,j}| = |A_{j,j}| > \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|,$$

so $0 \notin D_j$ for each $j \in \{1, \ldots, n\}$. Hence, 0 is not an eigenvalue of T, so T is invertible. □

Solution 5D.23 Let $A = \text{Mat}(T)$ be the matrix of T with respect to a basis $v_1, \ldots, v_n$ for V. Let λ be an eigenvalue of T. By Exercise 5B28, we know that T and T' have the same minimal polynomial, so they have the same eigenvalues. Moreover, we know that the matrix of T' with respect to the dual basis is just A^t. Using Theorem 5.67 on T', there must be some $k \in \{1, \ldots, n\}$ such that

$$|\lambda - A^t_{k,k}| \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |A^t_{k,j}|.$$

Since $A^t_{k,k} = A_{k,k}$ and $A^t_{k,j} = A_{j,k}$, we have

$$|\lambda - A_{k,k}| \leq \sum_{\substack{j=1 \\ j \neq k}}^{n} |A_{j,k}|,$$

so λ lies in the kth Gershgorin disk with the changed definition. □

§5E

Solution 5E.1 Let e_1, e_2, e_3, e_4 be the standard basis of $\mathbb{F}^4$. Consider the operators $S, T \in \text{End}(\mathbb{F}^4)$ defined by

$$S(x, y, z, w) = (0, x, 0, 0) \quad \text{and} \quad T(x, y, z, w) = (0, 0, x, 0).$$

Then

$$ST(x, y, z, w) = S(0, 0, x, 0) = (0, 0, 0, 0),$$
$$TS(x, y, z, w) = T(0, x, 0, 0) = (0, 0, 0, 0),$$

so $ST = TS$. The subspace $U = \text{span}(e_1, e_2)$ is invariant under S, since $Se_1 = e_2 \in U$ and $Se_2 = 0 \in U$, but not under T, since $Te_1 = e_3 \notin U$. The subspace $W = \text{span}(e_1, e_3)$ is invariant under T, since $Te_1 = e_3 \in W$ and $Te_3 = 0 \in W$, but not under S, since $Se_1 = e_2 \notin W$. □

Solution 5E.2

Axler Note. *This exercise extends 5.76, which considers the case in which $\mathscr{C}$ contains only two elements. For this exercise, $\mathscr{C}$ may contain any number of elements, and $\mathscr{C}$ may even be an infinite set.*

Suppose there exists a basis of V with respect to which every element of $\mathscr{C}$ has a diagonal matrix. Then every pair of elements of $\mathscr{C}$ commutes, since diagonal matrices commute.

Conversely, suppose every pair of elements of $\mathscr{C}$ commutes. We proceed by induction on $\dim V$. If $\dim V = 0$, then $\varnothing$ is a basis, every element of $\mathscr{C}$ has empty matrix representation, and every pair of elements of $\mathscr{C}$ commutes. If $\dim V = 1$, then every operator on V is diagonalizable and has a diagonal matrix with respect to any basis of V. Now suppose $\dim V > 1$ and the statement holds for all vector spaces of dimension

less than dim V. If every element of $\mathscr{C}$ is a scalar multiple of the identity operator, then every element of $\mathscr{C}$ has a diagonal matrix with respect to any basis of V. Otherwise, there exists an operator $S \in \mathscr{C}$ that is not a scalar multiple of the identity operator. Decompose V into a direct sum of eigenspaces of S:

$$V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S),$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of S. Since every element of $\mathscr{C}$ commutes with S, each eigenspace $E(\lambda_i, S)$ is invariant under every element of $\mathscr{C}$. Since S is not a scalar multiple of the identity, $m \geq 2$, so $\dim E(\lambda_i, S) < \dim V$ for each i. For any $T \in \mathscr{C}$, Theorem 5.62 implies that the minimal polynomial of T is a product of distinct linear factors. Since the minimal polynomial of $T|_{E(\lambda_i,S)}$ divides the minimal polynomial of T, it follows that $T|_{E(\lambda_i,S)}$ is diagonalizable. Furthermore, any pair of elements of $\mathscr{C}$ commutes on $E(\lambda_i, S)$ since they commute on V and $E(\lambda_i, S)$ is invariant. By the induction hypothesis, for each i there exists a basis of $E(\lambda_i, S)$ with respect to which every restricted element of $\mathscr{C}$ has a diagonal matrix. Concatenating these bases gives a basis of V with respect to which every element of $\mathscr{C}$ has a diagonal matrix. □

Solution 5E.3

Axler Note. *See 5.18 for the special case $S = T$.*

We first show that $ST = TS$ implies $S^n T = TS^n$ for every nonnegative integer n by induction on n. The base case $n = 0$ is trivial since $S^0 T = IT = T = TI = TS^0$. Now suppose $S^n T = TS^n$ for some nonnegative integer n. Then

$$S^{n+1}T = S(S^n T) = S(TS^n) = (ST)S^n = (TS)S^n = T(S^{n+1}),$$

so $S^{n+1}T = TS^{n+1}$. This completes the induction, so $S^n T = TS^n$ for any nonnegative integer n. Using this result, it follows that $p(S)T = Tp(S)$ for any polynomial $p \in \mathbb{F}[X]$.

(a) Suppose $v \in \ker p(S)$. Then $p(S)v = 0$, so

$$p(S)(Tv) = Tp(S)v = T(0) = 0,$$

which shows that $Tv \in \ker p(S)$. Thus, $\ker p(S)$ is invariant under T.

(b) Suppose $v \in \operatorname{im} p(S)$. Then there exists $u \in V$ such that $p(S)u = v$, so

$$p(S)(Tu) = Tp(S)u = T(v),$$

which shows that $Tv \in \operatorname{im} p(S)$. Thus, $\operatorname{im} p(S)$ is invariant under T. □

Solution 5E.4 This statement is false. Consider the diagonal matrix

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

and the upper-triangular matrix

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

Then

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

but

$$BA = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

Since $AB \neq BA$, A and B do not commute. □

Solution 5E.5

Axler Note. *See 3.118 for the definition of the dual of an operator.*

Suppose $S, T \in \mathrm{End}(V)$ commute. Then for any $\phi \in V'$ and $v \in V$,

$$(S'T'\phi)(v) = (T'\phi)(Sv) = \phi(TSv) = \phi(STv) = (S'\phi)(Tv) = (T'S'\phi)(v),$$

so $S'T' = T'S'$.

Conversely, suppose $S', T' \in \mathrm{End}(V')$ commute. Applying the result we just proved to S' and T', we have that $(S')'(T')' = (T')'(S')'$, or equivalently $S''T'' = T''S''$. By Exercise 3F32, we know that S and T are isomorphic to S'' and T'', respectively, via the isomorphism Λ. Then for any $v \in V$,

$$\Lambda(STv) = S''T''\Lambda(v) = T''S''\Lambda(v) = \Lambda(TSv),$$

and injectivity of Λ implies that $ST = TS$. □

Solution 5E.6 Assume $V \neq \{0\}$. Since V is finite-dimensional and over $\mathbb{C}$, S has an eigenvalue $\alpha \in \mathbb{C}$. Let $W = \mathrm{im}(S - \alpha I)$. Since $\ker(S - \alpha I) \neq \{0\}$, the fundamental theorem of linear maps gives $W \neq V$, so V/W is a nonzero complex finite-dimensional space. Since $ST = TS$, the subspace W is invariant under T: for $w = (S - \alpha I)u \in W$,

$$Tw = T(S - \alpha I)u = (S - \alpha I)Tu \in W.$$

Thus, the quotient operator $T/W \in \mathrm{End}(V/W)$ is well-defined and has an eigenvalue $\lambda \in \mathbb{C}$.

Suppose for contradiction that $W + \mathrm{im}(T - \lambda I) = V$. For every $v \in V$, write $v = w + (T - \lambda I)u$ with $w \in W$ and $u \in V$. Then

$$v + W = (T - \lambda I)u + W = (T/W - \lambda I)(u + W),$$

so $T/W - \lambda I$ is surjective on V/W. Since V/W is finite-dimensional, $T/W - \lambda I$ is invertible, contradicting λ being an eigenvalue of T/W. □

Solution 5E.7 Since S is diagonalizable, we can decompose V into a direct sum of eigenspaces of S:

$$V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S),$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of S. Since $ST = TS$, each eigenspace $E(\lambda_i, S)$ is invariant under T. Since V is complex and each $E(\lambda_i, S)$ is invariant under T, there exists a basis of $E(\lambda_i, S)$ with respect to which $T|_{E(\lambda_i, S)}$ has an upper-triangular matrix by Theorem 5.44. Let β_i be such a basis of $E(\lambda_i, S)$, and let β be

the union of the β_i. Since $V = E(\lambda_1, S) \oplus \cdots \oplus E(\lambda_m, S)$, β is a basis of V. Moreover, each basis vector in β_i is an eigenvector of S, so $\mathrm{Mat}(S, \beta)$ is diagonal. Lastly, for each $v \in \beta_i$, $Tv \in E(\lambda_i, S) = \mathrm{span}(\beta_i)$. By choice of β_i, $\mathrm{Mat}(T|_{E(\lambda_i,S)}, \beta_i)$ is upper-triangular, so $\mathrm{Mat}(T, \beta)$ is upper-triangular. □

Solution 5E.8 Consider the basis $1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3$ of $\mathbb{R}^2[X]_{\leq 3}$. Enumerate the basis vectors as $v_1, \ldots, v_{10}$ in the order given. Then $D_x v_j \in \mathrm{span}(v_1, \ldots, v_{j-1})$ and $D_y v_j \in \mathrm{span}(v_1, \ldots, v_{j-1})$ for each $j = 1, \ldots, 10$, so $\mathrm{Mat}(D_x)$ and $\mathrm{Mat}(D_y)$ are upper-triangular with respect to this basis. □

Solution 5E.9

Axler Note. *This exercise extends 5.78 and 5.80, which consider the case in which $\mathscr{E}$ contains only two elements. For this exercise, $\mathscr{E}$ may contain any number of elements, and $\mathscr{E}$ may even be an infinite set.*

(a) Let $T \in \mathscr{E}$. We proceed by induction on $\dim V$. For the base case $\dim V = 1$, every nonzero vector in V is an eigenvector for every element of $\mathscr{E}$. Now suppose $\dim V > 1$ and the statement holds for all complex vector spaces of dimension less than $\dim V$. If $\mathscr{E} = \varnothing$, then any nonzero vector in V is an eigenvector for every element of $\mathscr{E}$. Since V is complex and finite-dimensional, T has an eigenvalue $\lambda \in \mathbb{C}$. By Exercise 5E3, $E(\lambda, T)$ is invariant under every element of $\mathscr{E}$.

If $E(\lambda, T) = V$, then $T = \lambda I$, so every nonzero vector is an eigenvector of T. If all elements of $\mathscr{E}$ are scalar multiples of the identity, then every nonzero vector is an eigenvector for every element of $\mathscr{E}$. Otherwise, there exists an operator $S \in \mathscr{E}$ that is not a scalar multiple of the identity. Let μ be an eigenvalue of S. Then $E(\mu, S)$ is a nonzero proper subspace of V that is invariant under every element of $\mathscr{E}$ by Exercise 5E3, so $\dim E(\mu, S) < \dim V$. The restrictions $\{R|_{E(\mu,S)} : R \in \mathscr{E}\}$ pairwise commute because for any $R_1, R_2 \in \mathscr{E}$,

$$R_1|_{E(\mu,S)}\, R_2|_{E(\mu,S)} = (R_1 R_2)|_{E(\mu,S)} = (R_2 R_1)|_{E(\mu,S)} = R_2|_{E(\mu,S)}\, R_1|_{E(\mu,S)}.$$

By the inductive hypothesis, there exists a nonzero $v \in E(\mu, S)$ that is an eigenvector for every restriction, hence for every element of $\mathscr{E}$.

(b) We proceed by induction on $\dim V$. For the base case $\dim V = 1$, every operator on V has an upper-triangular matrix with respect to any basis of V. Now suppose $\dim V > 1$ and the statement holds for all complex vector spaces of dimension less than $\dim V$. By part (a), there exists a nonzero $v_1 \in V$ that is an eigenvector for every element of $\mathscr{E}$. Let $U = \mathrm{span}(v_1)$. Since v_1 is an eigenvector for every element of $\mathscr{E}$, U is invariant under every element of $\mathscr{E}$, so the quotient operators $\{T/U : T \in \mathscr{E}\}$ on V/U are well-defined. For any $S, T \in \mathscr{E}$,

$$(S/U)(T/U) = (ST)/U = (TS)/U = (T/U)(S/U),$$

so the quotient operators pairwise commute. Since $\dim(V/U) = \dim V - 1$, the inductive hypothesis gives a basis $v_2 + U, \ldots, v_n + U$ of V/U with respect to which every T/U has an upper-triangular matrix. Then $v_1, v_2, \ldots, v_n$ is a basis of V. For any $T \in \mathscr{E}$, $Tv_1 \in \mathrm{span}(v_1)$ since v_1 is an eigenvector of T. For $k \geq 2$,

T/U being upper-triangular gives $(T/U)(v_k + U) \in \operatorname{span}(v_2 + U, \ldots, v_k + U)$, so $Tv_k \in \operatorname{span}(v_1, v_2, \ldots, v_k)$. Hence, every element of $\mathscr{C}$ has an upper-triangular matrix with respect to $v_1, \ldots, v_n$.

□

Solution 5E.10

Axler Note. *This exercise shows that 5.81 does not hold on real vector spaces.*

Consider the operators $S, T \in \operatorname{End}(\mathbb{R}^2)$ defined by

$$S(x, y) = (-y, x) \quad \text{and} \quad T(x, y) = (y, -x).$$

Since both are rotation operators, they commute. Neither S nor T has any real eigenvalues since no nonzero vector in $\mathbb{R}^2$ is a scalar multiple of its image under either operator. However, $S + T$ is the zero operator, which has eigenvalue 0, and $ST = I$, which has eigenvalue 1. □

Chapter 6

§6A

Solution 6A.1 By the additivity of inner products in both slots, we have

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\langle v_j, v_k\rangle = \left\langle \sum_{j=1}^{m} v_j, \sum_{k=1}^{m} v_k \right\rangle = \left\| \sum_{j=1}^{m} v_j \right\|^2 \geq 0.$$ □

Solution 6A.2 Define $\langle u, v\rangle_1 = \langle Su, Sv\rangle$ for all $u, v \in V$. From the linearity of S and the properties of the regular inner product $\langle \cdot, \cdot \rangle$, we see that $\langle \cdot, \cdot \rangle_1$ satisfies the properties:

- $\langle u, v\rangle_1 = \langle Su, Sv\rangle = \overline{\langle Sv, Su\rangle} = \overline{\langle v, u\rangle_1}$ for all $u, v \in V$.
- $\langle u + w, v\rangle_1 = \langle S(u + w), Sv\rangle = \langle Su + Sw, Sv\rangle = \langle Su, Sv\rangle + \langle Sw, Sv\rangle = \langle u, v\rangle_1 + \langle w, v\rangle_1$ for all $u, v, w \in V$.
- $\langle \alpha u, v\rangle_1 = \langle S(\alpha u), Sv\rangle = \langle \alpha Su, Sv\rangle = \alpha\langle Su, Sv\rangle = \alpha\langle u, v\rangle_1$ for all $u, v \in V$ and $\alpha \in \mathbb{F}$.
- $\langle u, u\rangle_1 = \langle Su, Su\rangle = \|Su\|^2 \geq 0$ for all $u \in V$.

The only property that may not hold is definiteness.

If S is injective, then $\langle u, u\rangle_1 = 0 \implies \|Su\|^2 = 0 \implies Su = 0 \implies u = 0$, so definiteness holds. Conversely, if $\langle \cdot, \cdot \rangle_1$ is an inner product, then definiteness holds, so $Su = 0 \implies \langle u, u\rangle_1 = 0 \implies u = 0$, so S is injective. □

Solution 6A.3

(a) It is not linear in the first slot, since

$$\langle (1,0) + (-1,0), (1,0) \rangle = \langle (0,0), (1,0) \rangle = |0 \cdot 1| + |0 \cdot 0| = 0,$$

but

$$\langle (1,0), (1,0) \rangle + \langle (-1,0), (1,0) \rangle = |1 \cdot 1| + |0 \cdot 0| + |-1 \cdot 1| + |0 \cdot 0| = 2.$$

(b) It is not positive definite, since $(0,1,0) \neq (0,0,0)$ but

$$\langle (0,1,0), (0,1,0) \rangle = 0 \cdot 0 + 0 \cdot 0 = 0. \qquad \square$$

Solution 6A.4 Suppose $v \in V$ is such that $(T - \sqrt{2}I)v = 0$. Then $Tv = \sqrt{2}v$, so $\|Tv\| = \sqrt{2}\|v\|$. But $\|Tv\| \leq \|v\|$, so $\sqrt{2}\|v\| \leq \|v\|$, which implies $\|v\| = 0$, so $v = 0$. Thus, $T - \sqrt{2}I$ is injective. $\square$

Solution 6A.5

(a) We have

$$\begin{aligned} \langle u + v, u - v \rangle &= \langle u, u - v \rangle + \langle v, u - v \rangle \\ &= \langle u, u \rangle - \langle u, v \rangle + \langle v, u \rangle - \langle v, v \rangle \\ &= \|u\|^2 - \|v\|^2. \end{aligned}$$

(b) If $\|u\| = \|v\|$, then $\langle u + v, u - v \rangle = \|u\|^2 - \|v\|^2 = 0$, so $u + v$ is orthogonal to $u - v$.

(c) Let u and v be the vectors corresponding to two adjacent sides of a rhombus. Then $\|u\| = \|v\|$, so $u + v$ is orthogonal to $u - v$, where $u + v$ and $u - v$ represent the diagonals of the rhombus.

$\square$

Solution 6A.6 Suppose $\langle u, v \rangle = 0$. Then for all $a \in \mathbb{F}$, we have $\langle u, av \rangle = 0$. By the Pythagorean theorem, we have

$$\|u + av\|^2 = \|u\|^2 + \|av\|^2 \geq \|u\|^2,$$

Conversely, suppose that $\|u\| \leq \|u + av\|$ for all $a \in \mathbb{F}$. If $v = 0$, then $\langle u, v \rangle = 0$. Otherwise, consider $v \neq 0$, and let $a = -\frac{\langle u, v \rangle}{\|v\|^2}$. We have $\langle u + av, v \rangle = 0$, so $\langle u + av, av \rangle = 0$. Then by the Pythagorean theorem, we have

$$\|u\|^2 = \|(u + av) - av\|^2 = \|u + av\|^2 + \|av\|^2.$$

Note that $\|u\| \leq \|u + av\|$ implies $\|u\|^2 \leq \|u + av\|^2$. Hence, $\|av\|^2 \leq 0$, which implies $a = 0$ and $\langle u, v \rangle = 0$. $\square$

Solution 6A.7 For all $a, b \in \mathbb{R}$, we have

$$\begin{aligned}\|au + bv\|^2 = \|bu + av\|^2 &\iff \langle au + bv, au + bv\rangle = \langle bu + av, bu + av\rangle \\ &\iff a^2\|u\|^2 + 2ab\,\mathrm{Re}\langle u, v\rangle + b^2\|v\|^2 = b^2\|u\|^2 + 2ab\,\mathrm{Re}\langle u, v\rangle + a^2\|v\|^2 \\ &\iff (a^2 - b^2)(\|u\|^2 - \|v\|^2) = 0.\end{aligned}$$

Since there exist $a, b \in \mathbb{R}$ such that $a^2 - b^2 \neq 0$, we have $\|au + bv\| = \|bu + av\|$ for all $a, b \in \mathbb{R}$ if and only if $\|u\|^2 - \|v\|^2 = 0$, which is equivalent to $\|u\| = \|v\|$. □

Solution 6A.8 Let $u = (a, b, c, x, y)$ and $v = (1, 1, 1, 4, 9)$. Then $\|u\|^2 = a^2 + b^2 + c^2 + x^2 + y^2 \leq 1$ and $\|v\|^2 = 1^2 + 1^2 + 1^2 + 4^2 + 9^2 = 100$. By the Cauchy–Schwarz inequality, we have

$$a + b + c + 4x + 9y = \langle u, v\rangle \leq |\langle u, v\rangle| \leq \|u\|\,\|v\| \leq \sqrt{100} = 10.$$ □

Solution 6A.9 We have

$$\|u - v\|^2 = \langle u - v, u - v\rangle = \langle u, u\rangle - \langle u, v\rangle - \langle v, u\rangle + \langle v, v\rangle = 1 - 1 - 1 + 1 = 0,$$

where $\langle v, u\rangle = \overline{\langle u, v\rangle} = \overline{1} = 1$. Therefore, $\|u - v\| = 0$, which implies $u - v = 0$, so $u = v$. □

Solution 6A.10 Consider the vectors $(\|u\|, \sqrt{1 - \|u\|^2})$ and $(\|v\|, \sqrt{1 - \|v\|^2})$ in $\mathbb{R}^2$. By the Cauchy–Schwarz inequality, we have

$$\|u\|\,\|v\| + \sqrt{1 - \|u\|^2}\sqrt{1 - \|v\|^2} \leq \sqrt{\|u\|^2 + 1 - \|u\|^2}\sqrt{\|v\|^2 + 1 - \|v\|^2} = 1.$$

By the Cauchy–Schwarz inequality, we also have $|\langle u, v\rangle| \leq \|u\|\,\|v\|$. Therefore,

$$\sqrt{1 - \|u\|^2}\sqrt{1 - \|v\|^2} \leq 1 - \|u\|\,\|v\| \leq 1 - |\langle u, v\rangle|.$$ □

Solution 6A.11 Let $u = a(1, 3) = (a, 3a)$ for some $a \in \mathbb{R}$. Then $v = (1, 2) - u = (1 - a, 2 - 3a)$. Since v is orthogonal to $(1, 3)$, we have

$$\langle v, (1, 3)\rangle = (1 - a) + 3(2 - 3a) = 7 - 10a = 0,$$

which implies $a = \frac{7}{10}$. Therefore, $u = \left(\frac{7}{10}, \frac{21}{10}\right)$ and $v = \left(\frac{3}{10}, -\frac{1}{10}\right)$. □

Solution 6A.12

(a) Consider the vectors in $\mathbb{R}^4$:

$$\begin{aligned}u &= \left(\sqrt{a}, \sqrt{b}, \sqrt{c}, \sqrt{d}\right) \implies \|u\|^2 = a + b + c + d, \\ v &= \left(\frac{1}{\sqrt{a}}, \frac{1}{\sqrt{b}}, \frac{1}{\sqrt{c}}, \frac{1}{\sqrt{d}}\right) \implies \|v\|^2 = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}.\end{aligned}$$

Note that

$$\langle u, v\rangle = \sqrt{a}\cdot\frac{1}{\sqrt{a}} + \sqrt{b}\cdot\frac{1}{\sqrt{b}} + \sqrt{c}\cdot\frac{1}{\sqrt{c}} + \sqrt{d}\cdot\frac{1}{\sqrt{d}} = 4.$$

By the Cauchy–Schwarz inequality, we have $|\langle u, v\rangle| \le \|u\|\,\|v\|$, so

$$16 = |\langle u, v\rangle|^2 \le \|u\|^2\,\|v\|^2 = (a+b+c+d)\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}\right).$$

(b) By the equality condition of the Cauchy–Schwarz inequality, we have equality if and only if u is a scalar multiple of v. If $u = \alpha v$ for some $\alpha \in \mathbb{R}$, then

$$\begin{cases} \sqrt{a} = \alpha\frac{1}{\sqrt{a}} \implies a = \alpha, \\ \sqrt{b} = \alpha\frac{1}{\sqrt{b}} \implies b = \alpha, \\ \sqrt{c} = \alpha\frac{1}{\sqrt{c}} \implies c = \alpha, \\ \sqrt{d} = \alpha\frac{1}{\sqrt{d}} \implies d = \alpha, \end{cases}$$

so $a = b = c = d$.

□

Solution 6A.13

Manual Remark. *This is known as the QM–AM inequality, which is a special case of the power mean inequality.*

Let $u = (a_1, \ldots, a_n)$ and $v = (1, \ldots, 1) \in \mathbb{R}^n$. Then $\|u\|^2 = a_1^2 + \cdots + a_n^2$ and $\|v\|^2 = n$. By the Cauchy–Schwarz inequality, we have

$$|a_1 + \cdots + a_n|^2 = |\langle u, v\rangle|^2 \le \|u\|^2\,\|v\|^2 = n(a_1^2 + \cdots + a_n^2),$$

so

$$\left(\frac{a_1 + \cdots + a_n}{n}\right)^2 = \frac{|a_1 + \cdots + a_n|^2}{n^2} \le \frac{n(a_1^2 + \cdots + a_n^2)}{n^2} = \frac{a_1^2 + \cdots + a_n^2}{n}. \qquad \square$$

Solution 6A.14 Let $u \in V$ with $\|u\| = 1$. Then

$$\begin{aligned} \|v - u\|^2 &= \|v\|^2 - 2\,\mathrm{Re}\langle v, u\rangle + \|u\|^2 \\ &\ge \|v\|^2 - 2|\langle v, u\rangle| + 1 \\ &\ge \|v\|^2 - 2\|v\|\,\|u\| + 1 \\ &= \|v\|^2 - 2\|v\| + 1 \\ &= (\|v\| - 1)^2 \end{aligned}$$

Equality holds if and only if $\mathrm{Re}\langle v, u\rangle = |\langle v, u\rangle| = \|v\|$. The second equality implies $v = \alpha u$ for some $\alpha \in \mathbb{F}$. Since $\alpha = \langle v, u\rangle$ and $\mathrm{Re}\langle v, u\rangle = |\langle v, u\rangle|$, we have $\alpha = \|v\| > 0$. Therefore, $u = v/\|v\|$. □

Solution 6A.15 Let θ be the angle between u and v. By the law of cosines, we have

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\|\cos\theta.$$

On the other hand, we have

$$\|u - v\|^2 = \langle u - v, u - v\rangle = \langle u, u\rangle - 2\langle u, v\rangle + \langle v, v\rangle = \|u\|^2 + \|v\|^2 - 2\langle u, v\rangle.$$

Therefore,

$$\|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\|\cos\theta = \|u\|^2 + \|v\|^2 - 2\langle u, v\rangle,$$

which implies $\langle u, v\rangle = \|u\|\,\|v\|\cos\theta$. □

Solution 6A.16 The Cauchy–Schwarz inequality states that $|\langle x, y\rangle| \le \|x\|\,\|y\|$ for all $x, y \in \mathbb{R}^n$. Therefore, if x and y are nonzero vectors in $\mathbb{R}^n$, then

$$-1 \le -\frac{|\langle x, y\rangle|}{\|x\|\,\|y\|} \le \frac{\langle x, y\rangle}{\|x\|\,\|y\|} \le \frac{|\langle x, y\rangle|}{\|x\|\,\|y\|} \le 1.$$

Since the domain of arccos is $[-1, 1]$, the Cauchy–Schwarz inequality ensures that arccos is defined for the given expression. □

Solution 6A.17 Define the vectors

$$u = \left(\sqrt{1}a_1, \sqrt{2}a_2, \ldots, \sqrt{n}a_n\right), \quad v = \left(\frac{b_1}{\sqrt{1}}, \frac{b_2}{\sqrt{2}}, \ldots, \frac{b_n}{\sqrt{n}}\right) \in \mathbb{R}^n.$$

Then

$$\|u\|^2 = \sum_{k=1}^{n} k a_k^2, \quad \|v\|^2 = \sum_{k=1}^{n} \frac{b_k^2}{k}, \quad \langle u, v\rangle = \sum_{k=1}^{n} a_k b_k.$$

By the Cauchy–Schwarz inequality, we have

$$\left(\sum_{k=1}^{n} a_k b_k\right)^2 = |\langle u, v\rangle|^2 \le \|u\|^2\,\|v\|^2 = \left(\sum_{k=1}^{n} k a_k^2\right)\left(\sum_{k=1}^{n} \frac{b_k^2}{k}\right).$$ □

Solution 6A.18

(a) Consider the functions over $[1, \infty)$:

$$u(x) = x f(x), \quad v(x) = \frac{1}{x}.$$

Then

$$\langle u, v\rangle = \int_1^\infty u(x)v(x)\,dx = \int_1^\infty f(x)\,dx, \quad \|u\|^2 = \int_1^\infty (x f(x))^2\,dx, \quad \|v\|^2 = \int_1^\infty \frac{1}{x^2}\,dx = 1.$$

By the Cauchy–Schwarz inequality, we have

$$\left(\int_1^\infty f\right)^2 = |\langle u, v\rangle|^2 \le \|u\|^2\,\|v\|^2 = \int_1^\infty x^2 (f(x))^2\,dx.$$

(b) The equality condition of the Cauchy–Schwarz inequality states that we have equality if and only if u is a scalar multiple of v. Note that $f(x) \geq 0$ for all $x \in [1, \infty)$, so the scalar multiple must be nonnegative. Therefore, we have equality if and only if there exists a nonnegative constant c such that $xf(x) = c/x$ for all $x \in [1, \infty)$, which is equivalent to $f(x) = c/x^2$ for all $x \in [1, \infty)$. Moreover, both sides are finite as follows:

$$\int_1^\infty f(x)\,dx = \int_1^\infty \frac{c}{x^2}\,dx = c, \quad \int_1^\infty x^2(f(x))^2\,dx = \int_1^\infty x^2\left(\frac{c}{x^2}\right)^2\,dx = c^2.$$

□

Solution 6A.19 Let $u \in V$ be an eigenvector of T corresponding to the eigenvalue λ, so $Tu = \lambda u$. Let $A = \mathrm{Mat}(T)$ be the matrix of T with respect to the basis $v_1, \ldots, v_n$. From $Tu = \lambda u$, the coordinate vector $x \in \mathbb{F}^n$ of u with respect to the basis $v_1, \ldots, v_n$ satisfies $Ax = \lambda x$. Note that $x \neq 0$ since $u \neq 0$.
Applying the Cauchy–Schwarz inequality to the rows of A, we have

$$|\lambda|^2\|x\|^2 = \|Ax\|^2 = \sum_{j=1}^n \left|\sum_{k=1}^n A_{j,k}x_k\right|^2 \leq \left(\sum_{j=1}^n \sum_{k=1}^n |A_{j,k}|^2\right)\|x\|^2.$$

Dividing both sides by $\|x\|^2$ gives the desired inequality. □

Solution 6A.20 By the triangle inequality, we have

$$\|u\| = \|(u - v) + v\| \leq \|u - v\| + \|v\|,$$

which implies $\|u\| - \|v\| \leq \|u - v\|$. Similarly, we have

$$\|v\| = \|(v - u) + u\| \leq \|v - u\| + \|u\| = \|u - v\| + \|u\|,$$

which implies $\|v\| - \|u\| \leq \|u - v\|$. Therefore, we have

$$-\|u - v\| \leq \|u\| - \|v\| \leq \|u - v\|,$$

which is equivalent to $\big|\,\|u\| - \|v\|\,\big| \leq \|u - v\|$. □

Solution 6A.21 By the parallelogram equality,

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2 = 16 + 36 = 52,$$

so $\|v\|^2 = 26 - 9 = 17$, which implies $\|v\| = \sqrt{17}$. □

Solution 6A.22 By the triangle inequality, we have

$$\begin{aligned}\|u + v\|^2\|u - v\|^2 &\leq (\|u\|^2 + \|v\|^2)(\|u\|^2 + \|-v\|^2)\\ &= (\|u\|^2 + \|v\|^2)(\|u\|^2 + \|v\|^2)\\ &= (\|u\|^2 + \|v\|^2)^2.\end{aligned}$$

Taking the square root of both sides gives the desired inequality. □

Solution 6A.23

Manual Remark. *Note that* $\min\{A, B\}^2 \leq \frac{A^2+B^2}{2}$ *so long as* $A, B \geq 0$. *This follows from the fact that either* $A \leq B$ *or* $B \leq A$ *forces* $\min\{A, B\}^2 \leq A^2$ *and* $\min\{A, B\}^2 \leq B^2$.

Appendix. *An alternate proof is given at the end of this solution.*

We proceed by induction on m. For the base case $m = 1$, we have $\|v_1\| \leq 1 = \sqrt{1}$, so the statement holds.

Now suppose the statement holds for some $m \geq 1$. Let $v_1, \ldots, v_{m+1} \in V$ be such that $\|v_k\| \leq 1$ for each $k = 1, \ldots, m+1$. By the induction hypothesis, there exist $a_1, \ldots, a_m \in \{1, -1\}$ such that

$$\|a_1 v_1 + \cdots + a_m v_m\| \leq \sqrt{m}.$$

Let $u = a_1 v_1 + \cdots + a_m v_m$. Then

$$\|u + v_{m+1}\|^2 = \|u\|^2 + 2\,\mathrm{Re}\langle u, v_{m+1}\rangle + \|v_{m+1}\|^2,$$

and

$$\|u - v_{m+1}\|^2 = \|u\|^2 - 2\,\mathrm{Re}\langle u, v_{m+1}\rangle + \|v_{m+1}\|^2.$$

Adding these two equations gives

$$\|u + v_{m+1}\|^2 + \|u - v_{m+1}\|^2 = 2\|u\|^2 + 2\|v_{m+1}\|^2.$$

Since $\|v_{m+1}\|^2 \leq 1$, we have

$$2\|u\|^2 + 2\|v_{m+1}\|^2 \leq 2\|u\|^2 + 2.$$

Therefore,

$$\min\{\|u + v_{m+1}\|, \|u - v_{m+1}\|\}^2 \leq \frac{\|u + v_{m+1}\|^2 + \|u - v_{m+1}\|^2}{2} \leq \|u\|^2 + 1 \leq m + 1,$$

where the last inequality follows from the induction hypothesis. Taking the square root of both sides gives

$$\min\{\|u + v_{m+1}\|, \|u - v_{m+1}\|\} \leq \sqrt{m+1}.$$

Therefore, there exists $a_{m+1} \in \{1, -1\}$ such that

$$\|a_1 v_1 + \cdots + a_m v_m + a_{m+1} v_{m+1}\| = \|u + a_{m+1} v_{m+1}\| \leq \sqrt{m+1}. \qquad \square$$

Alternate proof. Coolempire93. Note that if the average value M of the squared norm $\|a_1 v_1 + \cdots + a_m v_m\|^2$ over all possible assignments of $a_1, \ldots, a_m \in \{1, -1\}$ is less than or equal to m, then there must exist some assignment $a_1, \ldots, a_m \in \{1, -1\}$ such that $\|a_1 v_1 + \cdots + a_m v_m\|^2 \leq m$. Taking the square root of both sides of this inequality would yield our desired inequality. By Exercise 6A13, this value is less than or equal to the average value M of the squared norm $\|a_1 v_1 + \cdots + a_m v_m\|^2$, so we may use M to bound our desired average.

Consider that there are 2^m possible assignments of $a_1, \ldots, a_m$, so

$$\begin{aligned}
M &= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \|a_1 v_1 + \cdots, a_m v_m\|^2 \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \left\| \sum_{k=1}^{m} a_k v_k \right\|^2 \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \left\langle \sum_{k=1}^{m} a_k v_k, \sum_{k=1}^{m} a_k v_k \right\rangle \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \left\langle \sum_{i=1}^{m} a_i v_i, \sum_{j=1}^{m} a_j v_j \right\rangle \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \sum_{i=1}^{m} \sum_{j=1}^{m} \langle a_i v_i, a_j v_j \rangle \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \langle v_i, v_j \rangle,
\end{aligned}$$

with the last steps being by additivity in both slots and homogeneity ($\overline{a_j} = a_j$ since a_j is real). Consider $a_i a_j \langle v_i, v_j \rangle$ when $i \neq j$. Because we are summing over all possible assignments of $a_1, \ldots, a_m$, and there are exactly as many assignments with $a_i a_j \langle v_i, v_j \rangle = \langle v_i, v_j \rangle$ as there are assignments with $a_i a_j \langle v_i, v_j \rangle = -\langle v_i, v_j \rangle$, these cancel out. We may simplify to only consider terms where $i = j$:

$$\begin{aligned}
M &= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \langle v_i, v_j \rangle \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \sum_{k=1}^{m} a_k a_k \langle v_k, v_k \rangle \\
&= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \sum_{k=1}^{m} a_k^2 \|v_k\|^2
\end{aligned}$$

But $1^2 = 1$ and $(-1)^2 = 1$, so $a_k^2 = 1$:

$$\begin{aligned}
M &= \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} \sum_{k=1}^{m} \|v_k\|^2 \\
&= \sum_{k=1}^{m} \|v_k\|^2 \cdot \frac{1}{2^m} \sum_{a_1,\ldots,a_m \in \{1,-1\}} 1 \\
&= \sum_{k=1}^{m} \|v_k\|^2 \cdot \frac{1}{2^m} 2^m \\
&= \sum_{k=1}^{m} \|v_k\|^2 \\
&\leq m
\end{aligned}$$

because there are 2^m assignments of $a_1, \ldots, a_m \in \{1, -1\}$, and our original assumption had $\|v_k\| \leq 1$.

Therefore, since the average of $\|a_1 v_1 + \cdots + a_m v_m\|^2$ over all possible assignments of $a_1, \ldots, a_m \in \{1, -1\}$ is less than or equal to m, there is an assignment $a_1, \ldots, a_m \in \{1, -1\}$ such that $\|a_1 v_1 + \cdots + a_m v_m\|^2 \leq m$. Taking the square root of both sides yields that for the given assignment, $\|a_1 v_1 + \cdots + a_m v_m\| \leq \sqrt{m}$.

Solution 6A.24 If we define the norm as $\|(x, y)\| = \max\{|x|, |y|\}$ for all $(x, y) \in \mathbb{R}^2$, then take $u = (1, 0)$ and $v = (0, 1)$. Then

$$\|u + v\|^2 + \|u - v\|^2 = 1 + 1 = 2,$$

but

$$2\|u\|^2 + 2\|v\|^2 = 2 + 2 = 4,$$

contradicting the parallelogram equality. □

Solution 6A.25 If $p = 2$, then the standard inner product on $\mathbb{R}^2$ gives the associated norm

$$\|(x, y)\| = (|x|^2 + |y|^2)^{1/2}.$$

Now suppose there is an inner product on $\mathbb{R}^2$ such that the associated norm is given by

$$\|(x, y)\| = (|x|^p + |y|^p)^{1/p}$$

for all $(x, y) \in \mathbb{R}^2$. Then the parallelogram equality must hold for this norm. Let $u = (1, 0)$ and $v = (0, 1)$. Then

$$\|u + v\|^2 + \|u - v\|^2 = 2^{2/p} + 2^{2/p} = 2 \cdot 2^{2/p} = 2^{1+2/p},$$

and

$$2\|u\|^2 + 2\|v\|^2 = 4.$$

Therefore, we must have

$$2^{1+2/p} = 4,$$

which implies $1 + 2/p = 2$, so $p = 2$. □

Solution 6A.26 We have

$$\begin{aligned}\|u + v\|^2 - \|u - v\|^2 &= \langle u + v, u + v\rangle - \langle u - v, u - v\rangle \\ &= (\|u\|^2 + 2\langle u, v\rangle + \|v\|^2) - (\|u\|^2 - 2\langle u, v\rangle + \|v\|^2) \\ &= 4\langle u, v\rangle.\end{aligned}$$

Dividing both sides by 4 gives the desired equality. □

Solution 6A.27 We have

$$\begin{aligned}\|u+v\|^2-\|u-v\|^2 &= \langle u+v,u+v\rangle-\langle u-v,u-v\rangle\\ &= (\|u\|^2+2\operatorname{Re}\langle u,v\rangle+\|v\|^2)-(\|u\|^2-2\operatorname{Re}\langle u,v\rangle+\|v\|^2)\\ &= 4\operatorname{Re}\langle u,v\rangle,\end{aligned}$$

and

$$\begin{aligned}\|u+iv\|^2-\|u-iv\|^2 &= \langle u+iv,u+iv\rangle-\langle u-iv,u-iv\rangle\\ &= (\|u\|^2+2\operatorname{Re}\langle u,iv\rangle+\|v\|^2)-(\|u\|^2-2\operatorname{Re}\langle u,iv\rangle+\|v\|^2)\\ &= 4\operatorname{Re}\langle u,iv\rangle.\end{aligned}$$

Let $\langle u,v\rangle = a+bi$ for some $a,b\in\mathbb{R}$. Then

$$\langle u,iv\rangle = \bar{\imath}\langle u,v\rangle = -i(a+bi) = b-ai,$$

so $\operatorname{Re}\langle u,iv\rangle = \operatorname{Im}\langle u,v\rangle$. Therefore, we have

$$\|u+v\|^2-\|u-v\|^2+\|u+iv\|^2i-\|u-iv\|^2i = 4\operatorname{Re}\langle u,v\rangle+4\operatorname{Im}\langle u,v\rangle i = 4\langle u,v\rangle.$$

Dividing both sides by 4 gives the desired equality. □

Solution 6A.28

Manual Remark. *This is a highly non-trivial result known as the* Jordan–von Neumann theorem. *The proof is not easy since it requires density of* $\mathbb{Q}$ *in* $\mathbb{R}$ *to ensure homogeneity from the rational numbers extends to the real numbers via continuity.*

Suppose $\|\cdot\|$ is a norm on U satisfying the parallelogram equality. We first work in the case that $\mathbb{F}=\mathbb{R}$. Define

$$\langle u,v\rangle = \frac{\|u+v\|^2-\|u-v\|^2}{4}$$

for all $u,v\in U$. We now check that $\langle\cdot,\cdot\rangle$ is an inner product on U.

- **Positive-Definiteness**: For any $u\in U$, we have

$$\langle u,u\rangle = \frac{\|u+u\|^2-\|u-u\|^2}{4} = \frac{\|2u\|^2-0}{4} = \frac{4\|u\|^2}{4} = \|u\|^2\geq 0,$$

  since $\|u\|\geq 0$. This establishes that $\|u\| = \langle u,u\rangle^{1/2}$ for all $u\in U$. Moreover, $\langle u,u\rangle = 0$ if and only if $\|u\|^2=0$, which is equivalent to $u=0$.

- **Symmetry**: Note that $\|a\| = \|-a\|$ for any $a\in U$. For any $u,v\in U$, we have

$$\langle u,v\rangle = \frac{\|u+v\|^2-\|u-v\|^2}{4} = \frac{\|v+u\|^2-\|v-u\|^2}{4} = \langle v,u\rangle.$$

- **Additivity in the First Slot**: Let $x,y,z\in U$. Applying the parallelogram equality to the pairs $(x+z,y)$ and $(x-z,y)$ gives

$$\begin{aligned}\|(x+z)+y\|^2+\|(x+z)-y\|^2 &= 2\|x+z\|^2+2\|y\|^2,\\ \|(x-z)+y\|^2+\|(x-z)-y\|^2 &= 2\|x-z\|^2+2\|y\|^2.\end{aligned}$$

Subtracting the second equation from the first gives

$$\|(x+z)+y\|^2-\|(x-z)+y\|^2+\|(x+z)-y\|^2-\|(x-z)-y\|^2 = 2\|x+z\|^2-2\|x-z\|^2.$$

Rearranging terms inside the norms gives

$$\|(x+y)+z\|^2-\|(x+y)-z\|^2+\|(x-y)+z\|^2-\|(x-y)-z\|^2 = 2\|x+z\|^2-2\|x-z\|^2.$$

Dividing both sides by 4 gives the relation $\langle x+y,z\rangle + \langle x-y,z\rangle = 2\langle x,z\rangle$. Note that $x = y$ yields $\langle 2x,z\rangle = 2\langle x,z\rangle$. Then for any $u,v,w \in V$, define

$$x = \frac{u+v}{2}, \quad y = \frac{u-v}{2}$$

so that

$$\langle u+v,w\rangle = \langle 2x,w\rangle = 2\langle x,w\rangle = \langle x+y,w\rangle + \langle x-y,w\rangle = \langle u,w\rangle + \langle v,w\rangle.$$

- **Homogeneity in the First Slot:** Let $u,v \in U$ and $\alpha \in \mathbb{R}$. We show that $\langle \alpha u,v\rangle = \alpha\langle u,v\rangle$ for all $\alpha \in \mathbb{R}$. If $\alpha = 0$, then

$$\langle 0,v\rangle = \frac{\|0+v\|^2-\|0-v\|^2}{4} = \frac{\|v\|^2-\|v\|^2}{4} = 0 = 0\langle u,v\rangle.$$

We first consider when α is a positive integer. We proceed by induction on α. For the base case $\alpha = 1$, we have $\langle 1\cdot u,v\rangle = \langle u,v\rangle$. Now suppose the statement holds for some positive integer α. Then

$$\langle(\alpha+1)u,v\rangle = \langle \alpha u+u,v\rangle = \langle \alpha u,v\rangle + \langle u,v\rangle = \alpha\langle u,v\rangle + \langle u,v\rangle = (\alpha+1)\langle u,v\rangle.$$

Therefore, the statement holds for all positive integers α. Next, we consider when α is a negative integer. By direct computation, we have for $\alpha = -1$ that

$$\langle -u,v\rangle = \frac{\|-u+v\|^2-\|-u-v\|^2}{4} = \frac{\|v-u\|^2-\|-(u+v)\|^2}{4} = \frac{\|v-u\|^2-\|u+v\|^2}{4} = -\langle u,v\rangle.$$

We may now consider $\alpha = -\beta$ for some positive integer β. Then

$$\langle \alpha u,v\rangle = \langle -\beta u,v\rangle = \langle -(\beta u),v\rangle = -\langle \beta u,v\rangle = -\beta\langle u,v\rangle = \alpha\langle u,v\rangle.$$

We now consider when α is a rational number. Let $\alpha = p/q$ for some integers p and q with $q > 0$. Then

$$q\langle \alpha u,v\rangle = \langle q(\alpha u),v\rangle = \langle pu,v\rangle = p\langle u,v\rangle,$$

so $\langle \alpha u,v\rangle = \alpha\langle u,v\rangle$.

Finally, we consider when α is an arbitrary real number. By the density of the rational numbers in the real numbers, there exists a sequence $(\alpha_n)_{n=1}^{\infty}$ of rational numbers such that $\alpha_n \to \alpha$ as $n \to \infty$. By the reverse triangle inequality, we have

$$\big|\,\|\alpha_n u+v\| - \|\alpha u+v\|\,\big| \le \|(\alpha_n-\alpha)u\| = |\alpha_n-\alpha|\,\|u\|,$$

where $|\alpha_n-\alpha| \to 0$ as $n \to \infty$, so $\|\alpha_n u+v\| \to \|\alpha u+v\|$ as $n \to \infty$. Similarly, we have $\|\alpha_n u-v\| \to \|\alpha u-v\|$ as $n \to \infty$. Then $\langle \alpha_n u,v\rangle \to \langle \alpha u,v\rangle$ as $n \to \infty$. It follows that $\langle \alpha_n u,v\rangle = \alpha_n\langle u,v\rangle \to \alpha\langle u,v\rangle$ as $n \to \infty$, so $\langle \alpha u,v\rangle = \alpha\langle u,v\rangle$.

Now suppose $\mathbb{F} = \mathbb{C}$. Define

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4} + \frac{\|u+iv\|^2 - \|u-iv\|^2}{4} i$$

for all $u, v \in U$. We again check that $\langle \cdot, \cdot \rangle$ is an inner product on U.

- **Positive-Definiteness:** For any $u \in U$, $\mathrm{Re}\langle u, u\rangle$ follows from the real case. Considering $\mathrm{Im}\langle u, u\rangle$, we have that

$$\begin{aligned}\|u+iu\|^2 &= \|(1+i)u\|^2 = |1+i|^2\|u\|^2 = 2\|u\|^2,\\ \|u-iu\|^2 &= \|(1-i)u\|^2 = |1-i|^2\|u\|^2 = 2\|u\|^2,\end{aligned}$$

  so $\mathrm{Im}\langle u, u\rangle = 0$. Therefore, $\langle u, u\rangle = \mathrm{Re}\langle u, u\rangle + \mathrm{Im}\langle u, u\rangle i = \mathrm{Re}\langle u, u\rangle \geq 0$. This again establishes that $\|u\| = \langle u, u\rangle^{1/2}$ for all $u \in U$. Moreover, $\langle u, u\rangle = 0$ if and only if $\mathrm{Re}\langle u, u\rangle = 0$, which is equivalent to $u = 0$.

- **Conjugate Symmetry:** For any $u, v \in U$, $\mathrm{Re}\langle u, v\rangle$ follows from the real case. Moreover, note that $\|u+iv\| = \|i(v-iu)\| = |i|\,\|v-iu\| = \|v-iu\|$. Similarly, $\|u-iv\| = \|-i(v+iu)\| = |-i|\,\|v+iu\| = \|v+iu\|$. Thus, for $\mathrm{Im}\langle u, v\rangle$, we have

$$\begin{aligned}\mathrm{Im}\langle u, v\rangle &= \frac{\|u+iv\|^2 - \|u-iv\|^2}{4}\\ &= \frac{\|v-iu\|^2 - \|v+iu\|^2}{4}\\ &= -\mathrm{Im}\langle v, u\rangle.\end{aligned}$$

  Therefore, $\langle u, v\rangle = \mathrm{Re}\langle u, v\rangle + \mathrm{Im}\langle u, v\rangle i = \mathrm{Re}\langle v, u\rangle - \mathrm{Im}\langle v, u\rangle i = \overline{\langle v, u\rangle}$.

- **Additivity in the First Slot:** Let $u, v, w \in U$. Then $\mathrm{Re}\langle u+v, w\rangle$ follows from the real case. For $\mathrm{Im}\langle u+v, w\rangle$, the real-case additivity identity applied with w replaced by iw gives

$$\mathrm{Im}\langle u+v, w\rangle = \mathrm{Im}\langle u, w\rangle + \mathrm{Im}\langle v, w\rangle.$$

  Therefore,

$$\begin{aligned}\langle u+v, w\rangle &= \mathrm{Re}\langle u+v, w\rangle + \mathrm{Im}\langle u+v, w\rangle i\\ &= \mathrm{Re}\langle u, w\rangle + \mathrm{Re}\langle v, w\rangle + (\mathrm{Im}\langle u, w\rangle + \mathrm{Im}\langle v, w\rangle)i\\ &= (\mathrm{Re}\langle u, w\rangle + \mathrm{Im}\langle u, w\rangle i) + (\mathrm{Re}\langle v, w\rangle + \mathrm{Im}\langle v, w\rangle i)\\ &= \langle u, w\rangle + \langle v, w\rangle.\end{aligned}$$

- **Homogeneity in the First Slot:** Let $u, v \in U$ and $\alpha \in \mathbb{C}$. We first consider $\alpha = i$. Then

$$\begin{aligned}\langle iu, v\rangle &= \frac{\|iu+v\|^2 - \|iu-v\|^2}{4} + \frac{\|iu+iv\|^2 - \|iu-iv\|^2}{4} i\\ &= \frac{\|v-iu\|^2 - \|v+iu\|^2}{4} + \frac{\|i(u+v)\|^2 - \|i(u-v)\|^2}{4} i\\ &= -\mathrm{Im}\langle v, u\rangle + \mathrm{Re}\langle u, v\rangle i = i\langle u, v\rangle.\end{aligned}$$

  Now let $\alpha = a + bi$ for some $a, b \in \mathbb{R}$. From the real case, we have

$$\langle au, v\rangle = a\langle u, v\rangle \quad \text{and} \quad \langle ibu, v\rangle = b\langle iu, v\rangle = bi\langle u, v\rangle.$$

Therefore,

$$\langle \alpha u, v\rangle = \langle au + ibu, v\rangle = \langle au, v\rangle + \langle ibu, v\rangle = a\langle u, v\rangle + bi\langle u, v\rangle = \alpha\langle u, v\rangle. \quad \square$$

Solution 6A.29 Let $\langle \cdot, \cdot \rangle_k$ denote the inner product on V_k for each $k = 1, \ldots, m$. To see that $\langle \cdot, \cdot \rangle$ is an inner product on $V_1 \times \cdots \times V_m$, note that the properties of conjugate symmetry, additivity in the first slot, and homogeneity in the first slot follow from the corresponding properties of each $\langle \cdot, \cdot \rangle_k$.
For positive-definiteness, let $(u_1, \ldots, u_m) \in V_1 \times \cdots \times V_m$. Then

$$\langle (u_1, \ldots, u_m), (u_1, \ldots, u_m)\rangle = \langle u_1, u_1\rangle_1 + \cdots + \langle u_m, u_m\rangle_m \geq 0.$$

Moreover, $\langle (u_1, \ldots, u_m), (u_1, \ldots, u_m)\rangle = 0$ if and only if $\langle u_k, u_k\rangle_k = 0$ for each $k = 1, \ldots, m$, which is equivalent to $u_k = 0$ for each $k = 1, \ldots, m$, so $(u_1, \ldots, u_m) = 0$. $\square$

Solution 6A.30

Axler Hint. *See Exercise 1B8 for the definition of the complexification* $V_{\mathbb{C}}$.

(a) We first check that $\langle \cdot, \cdot \rangle_{\mathbb{C}}$ is an inner product on $V_{\mathbb{C}}$.

- **Positive-Definiteness:** For any $u + iv \in V_{\mathbb{C}}$, we have

$$\langle u + iv, u + iv\rangle_{\mathbb{C}} = \langle u, u\rangle + \langle v, v\rangle + (\langle v, u\rangle - \langle u, v\rangle)i = \|u\|^2 + \|v\|^2 \geq 0,$$

since $\|u\|, \|v\| \geq 0$. Moreover, $\langle u+iv, u+iv\rangle_{\mathbb{C}} = 0$ if and only if $\|u\|^2+\|v\|^2 = 0$, which is equivalent to $u = v = 0$, so $u + iv = 0$.

- **Conjugate Symmetry:** For any $u + iv, w + ix \in V_{\mathbb{C}}$, we have

$$\begin{aligned}\langle w + ix, u + iv\rangle_{\mathbb{C}} &= \langle w, u\rangle + \langle x, v\rangle + (\langle x, u\rangle - \langle w, v\rangle)i \\ &= \overline{\langle u, w\rangle} + \overline{\langle v, x\rangle} + (\overline{\langle u, x\rangle} - \overline{\langle v, w\rangle})i \\ &= \overline{\langle u, w\rangle + \langle v, x\rangle + (\langle v, w\rangle - \langle u, x\rangle)i} = \overline{\langle u + iv, w + ix\rangle_{\mathbb{C}}}.\end{aligned}$$

- **Additivity in the First Slot:** Let $u + iv, w + ix, y + iz \in V_{\mathbb{C}}$. Then

$$\begin{aligned}\langle (u + iv) + (y + iz), w + ix\rangle_{\mathbb{C}} &= \langle (u + y) + i(v + z), w + ix\rangle_{\mathbb{C}} \\ &= \langle u + y, w\rangle + \langle v + z, x\rangle + (\langle v + z, w\rangle - \langle u + y, x\rangle)i\end{aligned}$$

Expanding the inner products gives

$$\begin{aligned}&(\langle u, w\rangle + \langle y, w\rangle) + (\langle v, x\rangle + \langle z, x\rangle) + ((\langle v, w\rangle + \langle z, w\rangle) - (\langle u, x\rangle + \langle y, x\rangle))i \\ &= (\langle u, w\rangle + \langle v, x\rangle + (\langle v, w\rangle - \langle u, x\rangle)i) + (\langle y, w\rangle + \langle z, x\rangle + (\langle z, w\rangle - \langle y, x\rangle)i) \\ &= \langle u + iv, w + ix\rangle_{\mathbb{C}} + \langle y + iz, w + ix\rangle_{\mathbb{C}}.\end{aligned}$$

- **Homogeneity in the First Slot:** Let $u + iv, w + ix \in V_{\mathbb{C}}$ and $\lambda \in \mathbb{R}$. Then

$$\begin{aligned}\langle \lambda(u + iv), w + ix\rangle_{\mathbb{C}} &= \langle (\lambda u) + i(\lambda v), w + ix\rangle_{\mathbb{C}} \\ &= \langle \lambda u, w\rangle + \langle \lambda v, x\rangle + (\langle \lambda v, w\rangle - \langle \lambda u, x\rangle)i \\ &= \lambda\langle u, w\rangle + \lambda\langle v, x\rangle + (\lambda\langle v, w\rangle - \lambda\langle u, x\rangle)i \\ &= \lambda(\langle u, w\rangle + \langle v, x\rangle + (\langle v, w\rangle - \langle u, x\rangle)i) = \lambda\langle u + iv, w + ix\rangle_{\mathbb{C}}.\end{aligned}$$

Moreover, note that $i(u+iv) = -v+iu$. Then

$$\begin{aligned}\langle i(u+iv), w+ix\rangle_{\mathbb{C}} &= \langle -v+iu, w+ix\rangle_{\mathbb{C}}\\ &= \langle -v, w\rangle + \langle u, x\rangle + (\langle u, w\rangle - \langle -v, x\rangle)i\\ &= -\langle v, w\rangle + \langle u, x\rangle + (\langle u, w\rangle + \langle v, x\rangle)i\\ &= i(\langle u, w\rangle + \langle v, x\rangle + (\langle v, w\rangle - \langle u, x\rangle)i) = i\langle u+iv, w+ix\rangle_{\mathbb{C}}.\end{aligned}$$

Hence, homogeneity in the first slot holds for all $\alpha \in \mathbb{C}$ by using additivity in the first slot and the fact that $\alpha = a + bi$ for some $a, b \in \mathbb{R}$.

Hence, $\langle \cdot, \cdot \rangle_{\mathbb{C}}$ is an inner product on $V_{\mathbb{C}}$, so $V_{\mathbb{C}}$ is a complex inner product space.

(b) We have

$$\langle u, v\rangle_{\mathbb{C}} = \langle u+i0, v+i0\rangle_{\mathbb{C}} = \langle u, v\rangle + \langle 0, 0\rangle + (\langle 0, v\rangle - \langle u, 0\rangle)i = \langle u, v\rangle,$$

and

$$\|u+iv\|_{\mathbb{C}}^2 = \langle u+iv, u+iv\rangle_{\mathbb{C}} = \langle u, u\rangle + \langle v, v\rangle + (\langle v, u\rangle - \langle u, v\rangle)i = \|u\|^2 + \|v\|^2.$$

□

Solution 6A.31 Set $a = w-u$ and $b = w-v$. Then the left-hand side is

$$\|w - \tfrac{1}{2}(u+v)\|^2 = \|\tfrac{1}{2}(a+b)\|^2 = \frac{1}{4}\|a+b\|^2.$$

Moreover, the right-hand side is

$$\frac{\|w-u\|^2 + \|w-v\|^2}{2} - \frac{\|u-v\|^2}{4} = \frac{\|a\|^2 + \|b\|^2}{2} - \frac{\|a-b\|^2}{4} = \frac{2\|a\|^2 + 2\|b\|^2 - \|a-b\|^2}{4}.$$

From the parallelogram equality, we have

$$\|a+b\|^2 + \|a-b\|^2 = 2\|a\|^2 + 2\|b\|^2 \implies \|a+b\|^2 = 2\|a\|^2 + 2\|b\|^2 - \|a-b\|^2,$$

so division by 4 gives our desired equality. □

Solution 6A.32 Suppose $u, v \in E$ are both closest points to w. Then $\|w-u\| = \|w-v\|$. By the property of E, we have $\frac{1}{2}(u+v) \in E$. By Exercise 6A31, we have

$$\|w - \tfrac{1}{2}(u+v)\|^2 = \frac{\|w-u\|^2 + \|w-v\|^2}{2} - \frac{\|u-v\|^2}{4} = \|w-u\|^2 - \frac{\|u-v\|^2}{4}.$$

Since $\frac{1}{2}(u+v) \in E$, we have

$$\|w-u\|^2 \le \|w - \tfrac{1}{2}(u+v)\|^2 = \|w-u\|^2 - \frac{\|u-v\|^2}{4},$$

so $\|u-v\|^2 = 0$, which implies that $u = v$. □

Solution 6A.33

Axler Note. *A function $f : \mathbb{R} \to \mathbb{R}^n$ is called differentiable if there exist differentiable functions $f_1, \ldots, f_n$ from $\mathbb{R}$ to $\mathbb{R}$ such that $f(t) = (f_1(t), \ldots, f_n(t))$ for each $t \in \mathbb{R}$. Furthermore, for each $t \in \mathbb{R}$, the derivative $f'(t) \in \mathbb{R}^n$ is defined by $f'(t) = (f_1'(t), \ldots, f_n'(t))$.*

(a) We have

$$\langle f(t), g(t)\rangle' = \left(\sum_{k=1}^{n} f_k(t)g_k(t)\right)' = \sum_{k=1}^{n}\big(f_k'(t)g_k(t) + f_k(t)g_k'(t)\big) = \langle f'(t), g(t)\rangle + \langle f(t), g'(t)\rangle.$$

(b) Suppose $\|f(t)\| = c$ for every $t \in \mathbb{R}$. Then $\langle f(t), f(t)\rangle = c^2$ for every $t \in \mathbb{R}$. Differentiating both sides with respect to t, we get

$$\langle f'(t), f(t)\rangle + \langle f(t), f'(t)\rangle = 2\langle f'(t), f(t)\rangle = 0,$$

so $\langle f'(t), f(t)\rangle = 0$ for every $t \in \mathbb{R}$.

(c) Geometrically, the result in (b) says that the tangent vector to a curve lying on a sphere in $\mathbb{R}^n$ centered at the origin is orthogonal to the position vector of the point on the curve.

□

Solution 6A.34 Consider $\triangle ABC$ with $\|C - A\| = a$, $\|C - B\| = b$, and $\|B - A\| = c$. Let M be the midpoint of $\overline{AB}$, so $M = \frac{1}{2}(A + B)$. Then $d = \|C - M\|$. By Exercise 6A31, we have

$$\|C - M\|^2 = \frac{\|C - A\|^2 + \|C - B\|^2}{2} - \frac{\|A - B\|^2}{4} = \frac{a^2 + b^2}{2} - \frac{c^2}{4}.$$

Rearranging terms gives $a^2 + b^2 = \frac{1}{2}c^2 + 2d^2$. □

Solution 6A.35 Consider $T \in \mathrm{End}(\mathbb{R}^n[X]_{\le m})$ defined by $Tr = \Delta((1 - \|x\|^2)r)$, where $\deg q = m$. Note that T is well-defined, since multiplication by $1 - \|x\|^2$ increases the degree of a polynomial by 2, and Δ decreases the degree by at most 2. Moreover, linearity of T follows from both linearity of Δ and linearity of multiplication by $1 - \|x\|^2$.

We show that T is injective. Suppose $Tr = 0$ for some $r \in \mathbb{R}^n[X]_{\le m}$. Then $\Delta((1 - \|x\|^2)r) = 0$, so $(1 - \|x\|^2)r$ is harmonic. Since $(1 - \|x\|^2)r(x) = 0$ for all $x \in \mathbb{R}^n$ with $\|x\| = 1$, we have $(1 - \|x\|^2)r = 0$ by the fact about harmonic functions given in the exercise.

Now fix $y \in \mathbb{R}^n$. Evaluating at $x = ty$ for $t \in \mathbb{R}$, we have $(1 - t^2\|y\|^2)r(ty) = 0$ for all $t \in \mathbb{R}$. If $y = 0$, then $r(ty) = r(0) = 0$ for all $t \in \mathbb{R}$. Otherwise, assume $y \neq 0$. Since $1 - t^2\|y\|^2$ is a polynomial in t of degree at most 2, it has at most two roots. Therefore, $r(ty) = 0$ for all $t \in \mathbb{R}$ except for at most two values of t. Since $r(ty)$ is a polynomial in t, it must be identically zero, so $r(ty) = 0$ for all $t \in \mathbb{R}$. Since $y \in \mathbb{R}^n$ was arbitrary, we have $r = 0$, so T is injective. Moreover, finite-dimensionality of

$\mathbb{R}^n[X]_{\leq m}$ implies that T is surjective. Therefore, there exists $r \in \mathbb{R}^n[X]_{\leq m}$ such that $Tr = \Delta((1 - \|x\|^2)r) = -\Delta q$. We then have

$$\Delta(q + (1 - \|x\|^2)r) = \Delta q + \Delta((1 - \|x\|^2)r) = \Delta q - \Delta q = 0,$$

so $p = q + (1 - \|x\|^2)r$ is harmonic. □

§6B

Solution 6B.1 Let $j, k \in \{1, \ldots, m\}$ with $j \neq k$. Then for $a_j = 1$ and $a_i = 0$ for all $i \neq j, k$, we have

$$\|e_j + a_k e_k\|^2 = |1|^2 + |a_k|^2 = 1 + |a_k|^2$$

by the hypothesis. If $a_k = 0$, then $\langle e_j, e_j \rangle = \|e_j\|^2 = 1$ for all j. Otherwise, let $|a_k| = 1$ so that $\|e_j + a_k e_k\|^2 = 1 + |a_k|^2 = 2$.
If $\mathbb{F} = \mathbb{R}$, then

$$\langle e_j, e_k \rangle = \frac{1}{4}(\|e_j + e_k\|^2 - \|e_j - e_k\|^2) = \frac{1}{4}(2 - 2) = 0$$

by Exercise 6A26. If $\mathbb{F} = \mathbb{C}$, then

$$\langle e_j, e_k \rangle = \frac{1}{4}(\|e_j + e_k\|^2 - \|e_j - e_k\|^2 + \|e_j + i e_k\|^2 i - \|e_j - i e_k\|^2 i) = \frac{1}{4}(2 - 2 + 2i - 2i) = 0$$

by Exercise 6A27. Therefore, the list $e_1, \ldots, e_m$ is orthonormal. □

Solution 6B.2

(a) Let $v_1 = (\cos\theta, \sin\theta)$ and $v_2 = (-\sin\theta, \cos\theta)$. Then $\|v_1\|^2 = \cos^2\theta + \sin^2\theta = 1$ and $\|v_2\|^2 = \sin^2\theta + \cos^2\theta = 1$. Moreover,

$$\langle v_1, v_2 \rangle = -\cos\theta \sin\theta + \sin\theta \cos\theta = 0.$$

Since v_1, v_2 is an orthonormal list of two vectors in $\mathbb{R}^2$, it is an orthonormal basis of $\mathbb{R}^2$. Similarly, let $w_1 = (\cos\theta, \sin\theta)$ and $w_2 = (\sin\theta, -\cos\theta)$. Then $\|w_1\|^2 = \cos^2\theta + \sin^2\theta = 1$ and $\|w_2\|^2 = \sin^2\theta + \cos^2\theta = 1$. Again,

$$\langle w_1, w_2 \rangle = \cos\theta \sin\theta - \sin\theta \cos\theta = 0.$$

Since w_1, w_2 is an orthonormal list of two vectors in $\mathbb{R}^2$, it is an orthonormal basis of $\mathbb{R}^2$.

(b) Suppose u_1, u_2 is an orthonormal basis of $\mathbb{R}^2$. If $u_1 = (x_1, y_1)$, then $\|u_1\|^2 = x_1^2 + y_1^2 = 1$, so there exists $\theta \in \mathbb{R}$ such that $x_1 = \cos\theta$ and $y_1 = \sin\theta$. If $u_2 = (x_2, y_2)$, then $\|u_2\|^2 = x_2^2 + y_2^2 = 1$ and $\langle u_1, u_2 \rangle = x_1 x_2 + y_1 y_2 = 0$. We now solve the system of equations

$$\begin{cases} x_2^2 + y_2^2 = 1, \\ \cos\theta\, x_2 + \sin\theta\, y_2 = 0. \end{cases}$$

If $\sin\theta = 0$, then $\theta = 0$ or $\theta = \pi$, so $\cos\theta = \pm 1$. Then $x_2 = 0$, so $y_2^2 = 1$, and thus $y_2 = \pm 1$. Then $u_2 = (0, \pm 1)$, which is of the form given in (a).

Otherwise, $y_2 = -\frac{\cos\theta}{\sin\theta} x_2$, so

$$x_2^2 + \left(-\frac{\cos\theta}{\sin\theta} x_2\right)^2 = 1 \implies x_2^2 + \frac{\cos^2\theta}{\sin^2\theta} x_2^2 = 1 \implies x_2^2 \left(1 + \frac{\cos^2\theta}{\sin^2\theta}\right) = 1.$$

We then have $x_2^2 = \sin^2\theta$, so $x_2 = \pm\sin\theta$. Then $y_2 = -\frac{\cos\theta}{\sin\theta}(\pm\sin\theta) = \mp\cos\theta$. Thus, $u_2 = (\pm\sin\theta, \mp\cos\theta)$, which is of the form given in (a).

□

Solution 6B.3 Since $e_1, \ldots, e_m$ is an orthonormal basis for $\text{span}(e_1, \ldots, e_m)$, if $v \in \text{span}(e_1, \ldots, e_m)$, then we have $\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2$.
Conversely, suppose $\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2$. Let $w = v - (\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m)$. Note that $\langle w, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle = 0$ for each $k = 1, \ldots, m$, so w is orthogonal to each e_k. Then

$$\|v\|^2 = \left\| w + \sum_{k=1}^{m} \langle v, e_k\rangle e_k \right\|^2 = \|w\|^2 + \sum_{k=1}^{m} |\langle v, e_k\rangle|^2.$$

The assumption of $\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2$ gives $\|w\|^2 = 0$, so $w = 0$, and $v \in \text{span}(e_1, \ldots, e_m)$. □

Solution 6B.4 Define the vectors

$$f_0 = \frac{1}{\sqrt{2\pi}}, \quad f_k = \frac{\cos kx}{\sqrt{\pi}}, \quad g_k = \frac{\sin kx}{\sqrt{\pi}} \quad \text{for } k = 1, \ldots, n.$$

Then

$$\|f_0\|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} 1\, dx = 1.$$

Moreover, note the following for $k = 1, \ldots, n$,

$$\langle f_0, f_k\rangle = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} \cos kx\, dx = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\pi}} \left.\frac{\sin kx}{k}\right|_{-\pi}^{\pi} = 0,$$

and $\langle f_0, g_k\rangle = 0$ because $\sin kx$ is an odd function. Hence, f_0 is orthogonal to each of the other vectors in the list. Using the $\cos kx$ and $\sin kx$ computations in the above integrals, we may simplify the calculations of the norms of f_k and g_k to get

$$\|f_k\|^2 = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos^2 kx\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + \cos 2kx)\, dx = 1,$$
$$\|g_k\|^2 = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin^2 kx\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - \cos 2kx)\, dx = 1.$$

We now show that the remaining vectors are pairwise orthogonal. For $j, k \in \{1, \ldots, n\}$ with $j \neq k$, we have that $j - k$ and $j + k$ are nonzero integers. Note that $\sin mx = 0$

for all integers m when $x = \pm\pi$. Then

$$\begin{aligned}
\langle f_j, f_k\rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \cos jx \cos kx\, dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} (\cos(j-k)x + \cos(j+k)x)\, dx \\
&= \frac{1}{2\pi}\left(\frac{\sin(j-k)x}{j-k} + \frac{\sin(j+k)x}{j+k}\right)\Bigg|_{-\pi}^{\pi} = 0, \\
\langle g_j, g_k\rangle &= \frac{1}{\pi}\int_{-\pi}^{\pi} \sin jx \sin kx\, dx \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} (\cos(j-k)x - \cos(j+k)x)\, dx \\
&= \frac{1}{2\pi}\left(\frac{\sin(j-k)x}{j-k} - \frac{\sin(j+k)x}{j+k}\right)\Bigg|_{-\pi}^{\pi} = 0.
\end{aligned}$$

For $j,k = 1,\ldots,n$, note that $f_j g_k$ is a product of an even function and an odd function, so $\langle f_j, g_k\rangle = 0$. Thus, the list of vectors is orthonormal. □

Solution 6B.5 The list of vectors in Exercise 6B4 is an orthonormal list in $C[-\pi,\pi]$. Using the same naming conventions of f_0, f_k, g_k as in Exercise 6B4, we have that

$$\langle f, f_0\rangle = \frac{a_0}{\sqrt{2}}, \quad \langle f, f_k\rangle = a_k, \quad \langle f, g_k\rangle = b_k.$$

By Bessel's inequality, for every $n \geq 1$,

$$\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2) = |\langle f, f_0\rangle|^2 + \sum_{k=1}^{n}(|\langle f, f_k\rangle|^2 + |\langle f, g_k\rangle|^2) \leq \|f\|^2 = \int_{-\pi}^{\pi} f^2.$$

Since the right side is independent of n, taking $n \to \infty$ gives the result. □

Solution 6B.6

Axler Note. *This exercise states in (a) that an appropriately small perturbation of an orthonormal basis is a basis. Then (b) shows that the number* $1/\sqrt{n}$ *on the right side of the inequality in (a) cannot be improved upon.*

(a) Suppose $v_1,\ldots,v_n$ are linearly dependent. Then there exist scalars $a_1,\ldots,a_n$, not all zero, such that

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

Then

$$\begin{aligned}\sqrt{\sum_{k=1}^{n}|a_k|^2} &= \left\|\sum_{k=1}^{n}a_k e_k\right\| \\ &= \left\|\sum_{k=1}^{n}a_k(e_k - v_k)\right\| \\ &\le \sum_{k=1}^{n}|a_k|\|e_k - v_k\| \\ &< \frac{1}{\sqrt{n}}\sum_{k=1}^{n}|a_k| \\ &\le \frac{1}{\sqrt{n}}\sqrt{n}\sqrt{\sum_{k=1}^{n}|a_k|^2} \\ &= \sqrt{\sum_{k=1}^{n}|a_k|^2}.\end{aligned}$$

This results in $1 < 1$, which is a contradiction, so $v_1, \ldots, v_n$ must be linearly independent. Since $\dim V = n$, it follows that $v_1, \ldots, v_n$ is a basis of V.

(b) Consider the vector

$$u = \frac{1}{\sqrt{n}}(e_1 + \cdots + e_n), \quad \|u\|^2 = \frac{1}{n}\|e_1 + \cdots + e_n\|^2 = \frac{1}{n}(n) = 1.$$

For each $k = 1, \ldots, n$, define

$$v_k = e_k - \frac{1}{\sqrt{n}}u \implies \|e_k - v_k\| = \|\frac{1}{\sqrt{n}}u\| = \frac{1}{\sqrt{n}}\|u\| = \frac{1}{\sqrt{n}}.$$

Then

$$v_1 + \cdots + v_n = (e_1 - \frac{1}{\sqrt{n}}u) + \cdots + (e_n - \frac{1}{\sqrt{n}}u) = (e_1 + \cdots + e_n) - \frac{n}{\sqrt{n}}u = (e_1 + \cdots + e_n) - \sqrt{n}u = 0,$$

so $v_1, \ldots, v_n$ is linearly dependent.

□

Solution 6B.7 Using the Gram–Schmidt procedure, we can produce an orthonormal basis of $\mathbb{R}^3$ from the given basis. Let $v_1 = (1, 0, 0), v_2 = (1, 1, 1), v_3 = (1, 1, 2)$. Set $f_1 = v_1 = (1, 0, 0)$, so $\|f_1\|^2 = 1^2 + 0^2 + 0^2 = 1$. Then

$$f_2 = v_2 - \frac{\langle v_2, f_1\rangle}{\|f_1\|^2}f_1 = (1, 1, 1) - \frac{1}{1}(1, 0, 0) = (0, 1, 1)$$

with $\|f_2\|^2 = 0^2 + 1^2 + 1^2 = 2$. Finally,

$$f_3 = v_3 - \frac{\langle v_3, f_1\rangle}{\|f_1\|^2}f_1 - \frac{\langle v_3, f_2\rangle}{\|f_2\|^2}f_2 = (1, 1, 2) - (1, 0, 0) - \frac{3}{2}(0, 1, 1) = \left(0, -\frac{1}{2}, \frac{1}{2}\right)$$

with $\|f_3\|^2 = 0^2 + (-\frac{1}{2})^2 + (\frac{1}{2})^2 = \frac{1}{2}$. Then the orthonormal basis of $\mathbb{R}^3$ is

$$e_1 = \frac{f_1}{\|f_1\|} = (1,0,0), \quad e_2 = \frac{f_2}{\|f_2\|} = \frac{1}{\sqrt{2}}(0,1,1), \quad e_3 = \frac{f_3}{\|f_3\|} = \sqrt{2}\left(0,-\frac{1}{2},\frac{1}{2}\right) = \left(0,-\frac{\sqrt{2}}{2},\frac{\sqrt{2}}{2}\right).$$

By the Gram–Schmidt procedure, we know $\operatorname{span}(e_1,\dots,e_k) = \operatorname{span}(v_1,\dots,v_k)$ for each $k = 1,2,3$. Since $\operatorname{Mat}(T,(v_1,v_2,v_3))$ is upper-triangular, $\operatorname{span}(v_1,\dots,v_k)$ is invariant under T for each $k = 1,2,3$. Hence, $\operatorname{span}(e_1,\dots,e_k)$ is invariant under T for each $k = 1,2,3$, so $\operatorname{Mat}(T,(e_1,e_2,e_3))$ is upper-triangular. □

Solution 6B.8

(a) Let $v_1 = 1, v_2 = x, v_3 = x^2$. Then $f_1 = v_1 = 1$, so $\|f_1\|^2 = \int_0^1 1^2 = 1$. Next,

$$f_2 = v_2 - \frac{\langle v_2, f_1\rangle}{\|f_1\|^2} f_1 = x - \int_0^1 x \cdot 1 = x - \frac{1}{2}$$

with $\|f_2\|^2 = \int_0^1 (x-\frac{1}{2})^2\,dx = \frac{1}{12}$. Finally,

$$f_3 = v_3 - \frac{\langle v_3, f_1\rangle}{\|f_1\|^2} f_1 - \frac{\langle v_3, f_2\rangle}{\|f_2\|^2} f_2 = x^2 - \int_0^1 x^2 \cdot 1\,dx - 12\left(\int_0^1 x^2\left(x-\frac{1}{2}\right)dx\right)\left(x-\frac{1}{2}\right) = x^2 - x + \frac{1}{6}$$

with $\|f_3\|^2 = \int_0^1 (x^2 - x + \frac{1}{6})^2\,dx = \frac{1}{180}$. Then the orthonormal basis of $\mathbb{R}[X]_{\le 2}$ is

$$e_1 = \frac{f_1}{\|f_1\|} = 1, \quad e_2 = \frac{f_2}{\|f_2\|} = 2\sqrt{3}\left(x-\frac{1}{2}\right), \quad e_3 = \frac{f_3}{\|f_3\|} = 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right).$$

(b) The differentiation operator D on $\mathbb{R}[X]_{\le 2}$ has the following matrix with respect to the basis $1, x, x^2$:

$$\operatorname{Mat}(D,(1,x,x^2)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then

$$De_1 = D(1) = 0,$$
$$De_2 = D\left(2\sqrt{3}\left(x-\frac{1}{2}\right)\right) = 2\sqrt{3} = 2\sqrt{3}e_1,$$
$$De_3 = D\left(6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right)\right) = 6\sqrt{5}(2x-1) = 6\sqrt{5}\cdot\frac{1}{\sqrt{3}}e_2 = 2\sqrt{15}e_2$$

Hence, the matrix of D with respect to the orthonormal basis e_1, e_2, e_3 is

$$\operatorname{Mat}(D,(e_1,e_2,e_3)) = \begin{pmatrix} 0 & 2\sqrt{3} & 0 \\ 0 & 0 & 2\sqrt{15} \\ 0 & 0 & 0 \end{pmatrix},$$

which is upper triangular, as expected.

□

Solution 6B.9 By the Gram–Schmidt procedure, we have

$$e_k = \frac{f_k}{\|f_k\|}, \quad f_k = v_k - \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j.$$

From the linear independence of $v_1, \ldots, v_m$, $v_k \notin \operatorname{span}(v_1, \ldots, v_{k-1}) = \operatorname{span}(e_1, \ldots, e_{k-1})$, so $f_k \neq 0$ and $\|f_k\| > 0$. Then by orthogonality of f_k and e_j for $j < k$, we have

$$\langle v_k, e_k \rangle = \frac{1}{\|f_k\|} \langle v_k, f_k \rangle = \frac{1}{\|f_k\|} \left\langle f_k + \sum_{j=1}^{k-1} \langle v_k, e_j \rangle e_j, f_k \right\rangle = \frac{1}{\|f_k\|} \langle f_k, f_k \rangle = \|f_k\| > 0.$$

□

Solution 6B.10 Suppose $e_1, \ldots, e_m$ is an orthonormal list in V such that $\langle v_k, e_k \rangle > 0$ and $\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k)$ for each $k = 1, \ldots, m$. We induce on k to show that e_k agrees with the Gram–Schmidt output.

For $k = 1$, we have $\operatorname{span}(v_1) = \operatorname{span}(e_1)$ and $\|e_1\| = 1$, so $v_1 = \langle v_1, e_1 \rangle e_1$. Taking norms and using $\langle v_1, e_1 \rangle > 0$ gives $\langle v_1, e_1 \rangle = \|v_1\|$, so $e_1 = v_1 / \|v_1\|$, which is the Gram–Schmidt output.

Now suppose the claim holds for $e_1, \ldots, e_k$. Since $v_{k+1} \in \operatorname{span}(e_1, \ldots, e_{k+1})$ and $e_1, \ldots, e_{k+1}$ is orthonormal,

$$v_{k+1} = \sum_{j=1}^{k+1} \langle v_{k+1}, e_j \rangle e_j \implies \langle v_{k+1}, e_{k+1} \rangle e_{k+1} = v_{k+1} - \sum_{j=1}^{k} \langle v_{k+1}, e_j \rangle e_j.$$

By the induction hypothesis, the right-hand side is the Gram–Schmidt vector f_{k+1}. Taking norms and using $\langle v_{k+1}, e_{k+1} \rangle > 0$ gives $\langle v_{k+1}, e_{k+1} \rangle = \|f_{k+1}\|$, so $e_{k+1} = f_{k+1} / \|f_{k+1}\|$, which is the Gram–Schmidt output. □

Solution 6B.11 The map $p \mapsto p(\frac{1}{2})$ is a linear functional on $\mathbb{R}[X]_{\leq 2}$. By the Riesz representation theorem, there exists a unique $q \in \mathbb{R}[X]_{\leq 2}$ such that $p(\frac{1}{2}) = \langle p, q \rangle$ for every $p \in \mathbb{R}[X]_{\leq 2}$, given by

$$q = e_1(\tfrac{1}{2})\, e_1 + e_2(\tfrac{1}{2})\, e_2 + e_3(\tfrac{1}{2})\, e_3,$$

where e_1, e_2, e_3 is the orthonormal basis from Exercise 6B8. Evaluating e_1, e_2, e_3 at $\frac{1}{2}$ gives

$$e_1(\tfrac{1}{2}) = 1, \quad e_2(\tfrac{1}{2}) = 2\sqrt{3}\left(\frac{1}{2} - \frac{1}{2}\right) = 0, \quad e_3(\tfrac{1}{2}) = 6\sqrt{5}\left(\frac{1}{4} - \frac{1}{2} + \frac{1}{6}\right) = -\frac{\sqrt{5}}{2}.$$

Hence, the polynomial q is

$$q = e_1 - \frac{\sqrt{5}}{2} e_3 = 1 - 15\left(x^2 - x + \frac{1}{6}\right) = -15x^2 + 15x - \frac{3}{2}.$$

□

Solution 6B.12 By the Riesz representation theorem, there exists a unique $q \in \mathbb{R}[X]_{\leq 2}$ such that

$$\int_0^1 p(x)\cos(\pi x)\,dx = \langle p, q\rangle = \int_0^1 pq$$

for every $p \in \mathbb{R}[X]_{\leq 2}$, given by

$$q = \langle e_1, \cos(\pi x)\rangle e_1 + \langle e_2, \cos(\pi x)\rangle e_2 + \langle e_3, \cos(\pi x)\rangle e_3,$$

where e_1, e_2, e_3 is the orthonormal basis from Exercise 6B8. Evaluating the inner products gives

$$\begin{aligned}
\langle e_1, \cos(\pi x)\rangle &= \int_0^1 e_1\cos(\pi x) = \int_0^1 \cos(\pi x) = 0,\\
\langle e_2, \cos(\pi x)\rangle &= \int_0^1 e_2\cos(\pi x) = \int_0^1 2\sqrt{3}\left(x - \frac{1}{2}\right)\cos(\pi x) = \frac{-4\sqrt{3}}{\pi^2},\\
\langle e_3, \cos(\pi x)\rangle &= \int_0^1 e_3\cos(\pi x) = \int_0^1 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right)\cos(\pi x) = 0.
\end{aligned}$$

Hence,

$$q = \frac{-4\sqrt{3}}{\pi^2}e_2 = \frac{-4\sqrt{3}}{\pi^2}\cdot 2\sqrt{3}\left(x - \frac{1}{2}\right) = \frac{-24}{\pi^2}\left(x - \frac{1}{2}\right). \qquad \square$$

Solution 6B.13 Let f_k, e_k be the vectors obtained from the Gram–Schmidt procedure, with the process continuing as long as $f_k \neq 0$. If $v_1, \ldots, v_m$ is linearly dependent, take the smallest k with $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$. Then $v_1, \ldots, v_{k-1}$ is linearly independent, so $e_1, \ldots, e_{k-1}$ is an orthonormal basis of that span. Hence,

$$v_k = \sum_{j=1}^{k-1}\langle v_k, e_j\rangle e_j \implies f_k = v_k - \sum_{j=1}^{k-1}\langle v_k, e_j\rangle e_j = 0.$$

Conversely, if the procedure gives $f_k = 0$ for some minimal k, then

$$v_k = \sum_{j=1}^{k-1}\langle v_k, e_j\rangle e_j \in \operatorname{span}(e_1, \ldots, e_{k-1}) = \operatorname{span}(v_1, \ldots, v_{k-1}),$$

so $v_1, \ldots, v_k$ is a linearly dependent list. $\square$

Solution 6B.14 By Exercise 6B10, there is a unique orthonormal list $e_1, \ldots, e_m$ such that $\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k)$ and $\langle v_k, e_k\rangle > 0$ for all k. Let $e'_1, \ldots, e'_m$ be any orthonormal list satisfying the same span condition. For each k, the vector e'_k lies in $\operatorname{span}(e_1, \ldots, e_k)$ and is orthogonal to $\operatorname{span}(e'_1, \ldots, e'_{k-1}) = \operatorname{span}(e_1, \ldots, e_{k-1})$. Now, write e'_k as

$$e'_k = \sum_{j=1}^{k} c_j e_j$$

for some scalars $c_1, \ldots, c_k$. Then $\langle e'_k, e_j\rangle = 0$ for $j < k$ gives $c_j = 0$ for $j < k$, so $e'_k = c_k e_k$. Since $\|e'_k\| = 1$ and V is real, we have $e'_k = \pm e_k$. These m sign choices are independent, giving exactly 2^m such lists. $\square$

Solution 6B.15 If $V = \{0\}$, then the result holds trivially. Otherwise, assume $V \neq \{0\}$, and let $e_1 \in V$ such that $\|e_1\|_2 = 1$. Let $c = \|e_1\|_1^2$. For any $u, v \in V$, $U = \operatorname{span}(u, v, e_1)$ is finite-dimensional. Extend e_1 to an orthonormal basis $e_1, \ldots, e_n$ of U with respect to $\langle \cdot, \cdot \rangle_2$. Since $\langle e_j, e_k \rangle_2 = 0$ for $j \neq k$, the same holds for $\langle \cdot, \cdot \rangle_1$.
For each $j > 1$, $\langle e_1 + e_j, e_1 - e_j \rangle_2 = 0$, so $\langle e_1 + e_j, e_1 - e_j \rangle_1 = 0$. Expanding gives

$$0 = \langle e_1 + e_j, e_1 - e_j \rangle_1 = \langle e_1, e_1 \rangle_1 - \langle e_j, e_j \rangle_1 = c - \langle e_j, e_j \rangle_1,$$

so $\langle e_j, e_j \rangle_1 = c > 0$ for all j. Write $u = a_1 e_1 + \cdots + a_n e_n$ and $v = b_1 e_1 + \cdots + b_n e_n$ for some scalars $a_1, \ldots, a_n, b_1, \ldots, b_n$. Then

$$\langle u, v \rangle_1 = \sum_{j=1}^{n} a_j \overline{b_j} \langle e_j, e_j \rangle_1 = c \sum_{j=1}^{n} a_j \overline{b_j} = c \langle u, v \rangle_2. \qquad \square$$

Solution 6B.16 If $V = \{0\}$, then the result holds trivially. Otherwise, assume $V \neq \{0\}$, and let $e_1, \ldots, e_n$ be an orthonormal basis of V with respect to $\langle \cdot, \cdot \rangle_2$. For any $v \in V$, write $v = a_1 e_1 + \cdots + a_n e_n$ for some scalars $a_1, \ldots, a_n$. Then

$$\|v\|_1 \leq \sum_{j=1}^{n} |a_j| \|e_j\|_1 \leq \left(\sum_{j=1}^{n} |a_j|^2 \right)^{1/2} \left(\sum_{j=1}^{n} \|e_j\|_1^2 \right)^{1/2} = \|v\|_2 \left(\sum_{j=1}^{n} \|e_j\|_1^2 \right)^{1/2}.$$

Taking $c = \left(\sum_{j=1}^{n} \|e_j\|_1^2 \right)^{1/2} > 0$ gives the desired inequality. $\square$

Solution 6B.17 By Schur's theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $\operatorname{Mat}(T)$ is upper-triangular. Since 1 is the only eigenvalue of T, the diagonal entries are all 1, so

$$Te_k = e_k + \sum_{i=1}^{k-1} a_{i,k} e_i$$

for each k. By orthonormality,

$$\|Te_k\|^2 = \left\| e_k + \sum_{i=1}^{k-1} a_{i,k} e_i \right\|^2 = 1 + \sum_{i=1}^{k-1} |a_{i,k}|^2.$$

Since $\|Te_k\| \leq \|e_k\| = 1$ by hypothesis, we have $\sum_{i=1}^{k-1} |a_{i,k}|^2 = 0$, forcing each $a_{i,k} = 0$. Thus, $Te_k = e_k$ for all k, so $T = I$. $\square$

Solution 6B.18 Let $U = \operatorname{span}(u_1, \ldots, u_m)$. Since $u_1, \ldots, u_m$ is a basis of U, there exists a unique linear functional $f : U \to \mathbb{F}$ satisfying $f(u_k) = 1$ for all $k \in \{1, \ldots, m\}$. Because U is a finite-dimensional inner product space (with the inner product inherited from V), the Riesz representation theorem provides a unique $v \in U$ such that $f(u) = \langle u, v \rangle$ for all $u \in U$. In particular, $\langle u_k, v \rangle = f(u_k) = 1$ for all k. $\square$

Solution 6B.19 For each $k \in \{1, \ldots, n\}$, since $v_1, \ldots, v_n$ is a basis of V, there is a unique linear functional $f_k : V \to \mathbb{F}$ satisfying $f_k(v_j) = \delta_{jk}$ for all $j \in \{1, \ldots, n\}$. By the Riesz representation theorem, there exists a unique $u_k \in V$ such that $f_k(v) = \langle v, u_k \rangle$ for all $v \in V$. In particular, $\langle v_j, u_k \rangle = f_k(v_j) = \delta_{jk}$. The list $u_1, \ldots, u_n$ is linearly independent because if $c_1 u_1 + \cdots + c_n u_n = 0$, then for each j,

$$0 = \left\langle v_j, \sum_{k=1}^{n} c_k u_k \right\rangle = \overline{c_j},$$

which implies $c_j = 0$ for all j. Hence, $u_1, \ldots, u_n$ is a basis of V with the desired properties. □

Solution 6B.20 By Exercise 5E9(b), there exists a basis $v_1, \ldots, v_n$ of V such that every element of $\mathscr{C}$ has an upper-triangular matrix with respect to this basis. Equivalently, every $T \in \mathscr{C}$ satisfies $T(\operatorname{span}(v_1, \ldots, v_k)) \subseteq \operatorname{span}(v_1, \ldots, v_k)$ for each $k \in \{1, \ldots, n\}$. Apply the Gram–Schmidt procedure to $v_1, \ldots, v_n$ to obtain an orthonormal basis $e_1, \ldots, e_n$ of V. By the key property of Gram–Schmidt,

$$\operatorname{span}(e_1, \ldots, e_k) = \operatorname{span}(v_1, \ldots, v_k)$$

for each k. Hence, every $T \in \mathscr{C}$ still satisfies $T(\operatorname{span}(e_1, \ldots, e_k)) \subseteq \operatorname{span}(e_1, \ldots, e_k)$ for each k, which means every element of $\mathscr{C}$ has an upper-triangular matrix with respect to the orthonormal basis $e_1, \ldots, e_n$. □

Solution 6B.21

Manual Remark. *This exercise is quite difficult with a long solution. A high level overview is this: Schur's theorem gives an orthonormal basis of V with respect to which* $\operatorname{Mat}(T)$ *is upper-triangular. This reduces the problem to showing each basis vector is contracted by T^m for sufficiently large m, which is done by induction on the index of the basis vector. The final step is to use the triangle inequality to show that T^m contracts every vector in V.*

If $V = \{0\}$, then the result holds trivially. Otherwise, assume $V \neq \{0\}$, and let $\dim V = n \geq 1$. By Schur's theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $\operatorname{Mat}(T)$ is upper-triangular whose diagonal entries are the eigenvalues of T. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, so $|\lambda_j| < 1$ for each j. Write $Te_j = \lambda_j e_j + v_j$ for each j, where $v_j \in \operatorname{span}(e_1, \ldots, e_{j-1})$ and $v_1 = 0$. Let

$$\lambda = \max_j |\lambda_j| < 1,$$

and choose α such that $\lambda < \alpha < 1$.

We prove by induction on j that there exists some $C_j > 0$ such that $\|T^k e_j\| \leq C_j \alpha^k$ for all $k \geq 0$. For the base case $j = 1$, we have $Te_1 = \lambda_1 e_1$, which gives

$$\|T^k e_1\| = |\lambda_1|^k \leq \lambda^k < \alpha^k,$$

so we can take $C_1 = 1$.

Assume the inductive hypothesis holds for all indices less than j. Consider

$$v_j = \sum_{i=1}^{j-1} a_i e_i.$$

By the inductive hypothesis, $\|T^k e_i\| \le C_i \alpha^k$ for all $i < j$, so

$$\|T^k v_j\| \le \sum_{i=1}^{j-1} |a_i| \, \|T^k e_i\| \le B\alpha^k, \quad B = \sum_{i=1}^{j-1} |a_i| C_i.$$

Repeatedly applying T where $Te_j = \lambda_j e_j + v_j$ gives

$$T^k e_j = \lambda_j^k e_j + \sum_{i=0}^{k-1} \lambda_j^{k-1-i} T^i v_j.$$

Taking norms and using the triangle inequality gives

$$\|T^k e_j\| \le \lambda^k + B \sum_{i=0}^{k-1} \lambda^{k-1-i} \alpha^i = \lambda^k + B \frac{\alpha^k - \lambda^k}{\alpha - \lambda} < \left(1 + \frac{B}{\alpha - \lambda}\right) \alpha^k,$$

so we can take $C_j = 1 + \frac{B}{\alpha - \lambda}$ to complete the inductive step.

Let $C = \max\{C_1, \ldots, C_n\}$. For any $v \in V$, write $v = a_1 e_1 + \cdots + a_n e_n$ for some scalars $a_1, \ldots, a_n$. Then $|a_j| \le \|v\|$ by orthonormality, so

$$\|T^k v\| \le \sum_{j=1}^{n} |a_j| \, \|T^k e_j\| \le nC\alpha^k \|v\|.$$

Choosing such m that $nC\alpha^m \le \varepsilon$ gives $\|T^m v\| \le \varepsilon \|v\|$ for every $v \in V$. □

Solution 6B.22 Suppose for contradiction that there exists $g \in C[-1, 1]$ such that $f(0) = \int_{-1}^{1} fg$ for every $f \in C[-1, 1]$. Take $f(x) = x^2 g(x)$, which is continuous. Then

$$0 = f(0) = \int_{-1}^{1} x^2 (g(x))^2 \, dx.$$

The integrand is continuous and nonnegative, so $x^2(g(x))^2 = 0$ for all $x \in [-1, 1]$. Hence, $g(x) = 0$ for all $x \in [-1, 1] \setminus \{0\}$, and continuity forces $g(0) = 0$, so $g \equiv 0$. But then $f(0) = 0$ for every $f \in C[-1, 1]$, which is false (take $f = 1$), a contradiction. Hence, no such g exists. □

Solution 6B.23

(a) We verify the metric axioms for d:

- **Non-negativity and Identity**: For all $u, v \in V$, $\|u - v\| \ge 0$ and $\|u - v\| = 0$ if and only if $u = v$.
- **Symmetry**: For all $u, v \in V$, we have

$$d(u, v) = \|u - v\| = \|-(v - u)\| = \|v - u\| = d(v, u).$$

- **Triangle Inequality**: For all $u, v, w \in V$, we have

$$d(u,w) = \|u-w\| = \|(u-v)+(v-w)\| \le \|u-v\|+\|v-w\| = d(u,v)+d(v,w).$$

Therefore, (V,d) is a metric space.

(b) If V is finite-dimensional, let $e_1, \dots, e_n$ be an orthonormal basis of V. A Cauchy sequence $(v_k)_{k\ge1}$ in V can be written as $v_k = a_{k,1}e_1 + \cdots + a_{k,n}e_n$ for scalars $a_{k,j}$. By the Cauchy–Schwarz inequality, for each j,

$$|a_{k,j} - a_{m,j}| = |\langle v_k - v_m, e_j\rangle| \le \|v_k - v_m\|,$$

so each coordinate sequence $(a_{k,j})_{k\ge1}$ is Cauchy in $\mathbb{F}$. Since $\mathbb{F}$ is complete, each $(a_{k,j})$ converges to some $a_j \in \mathbb{F}$. Define $v = a_1e_1 + \cdots + a_ne_n \in V$. Since $e_1, \dots, e_n$ is orthonormal,

$$\|v_k - v\|^2 = \sum_{j=1}^{n} |a_{k,j} - a_j|^2 \to 0,$$

so $v_k \to v$ in V, and (V,d) is complete.

(c) Let U be a finite-dimensional subspace of V, and let $(u_k)_{k\ge1}$ be a sequence in U converging to some $v \in V$. Since (u_k) converges, it is Cauchy. Restricting the inner product on V to U, we see that (u_k) is also a Cauchy sequence in U. By part (b), U is complete, so (u_k) converges to some $u \in U$. Since (u_k) converges to v in V and to u in U, we have $v = u \in U$. Therefore, U is closed in V.

□

§6C

Solution 6C.1 Suppose $v \in \{v_1, \dots, v_m\}^\perp$. Then $\langle v, v_k\rangle = 0$ for each $k = 1, \dots, m$. For any $u \in \operatorname{span}(v_1, \dots, v_m)$, we can write $u = a_1v_1 + \cdots + a_mv_m$ for some scalars $a_1, \dots, a_m$. Then

$$\langle v, u\rangle = \sum_{k=1}^{m} \overline{a_k}\langle v, v_k\rangle = 0,$$

so $v \in (\operatorname{span}(v_1, \dots, v_m))^\perp$.
Conversely, suppose $v \in (\operatorname{span}(v_1, \dots, v_m))^\perp$. Since $\{v_1, \dots, v_m\} \subseteq \operatorname{span}(v_1, \dots, v_m)$, Theorem 6.48(e) shows that $\operatorname{span}(v_1, \dots, v_m)^\perp \subseteq \{v_1, \dots, v_m\}^\perp$, so $v \in \{v_1, \dots, v_m\}^\perp$.

□

Solution 6C.2 Let $e_1, \dots, e_m, f_1, \dots, f_n$ be the result of applying the Gram–Schmidt procedure to the basis $u_1, \dots, u_m, v_1, \dots, v_n$ of V. By the properties of the Gram–Schmidt procedure, $\operatorname{span}(e_1, \dots, e_k) = \operatorname{span}(u_1, \dots, u_k)$ for each $k \le m$. In particular, $\operatorname{span}(e_1, \dots, e_m) = \operatorname{span}(u_1, \dots, u_m) = U$, so $e_1, \dots, e_m$ is an orthonormal basis of U.

Since each f_j is constructed to be orthogonal to all previous vectors in the Gram–Schmidt process, it follows that each f_j is orthogonal to every e_k for $k = 1, \ldots, m$. Hence, $f_1, \ldots, f_n \in U^\perp$. Additionally, $f_1, \ldots, f_n$ is an orthonormal list in $U^\perp$ because the Gram–Schmidt procedure produces orthonormal vectors. By Theorem 6.49, $V = U \oplus U^\perp$, so $\dim U + \dim U^\perp = \dim V$. Since $\dim U = m$ and $\dim V = m + n$, we have $\dim U^\perp = n$. Therefore, $f_1, \ldots, f_n$ is an orthonormal basis of $U^\perp$. □

Solution 6C.3 Let $v_1 = (1, 2, 3, -4)$ and $v_2 = (-5, 4, 3, 2)$. Applying the Gram–Schmidt procedure to v_1 and v_2, we first set $f_1 = v_1$ with $\|f_1\|^2 = 30$. Then

$$f_2 = v_2 - \frac{\langle v_2, f_1 \rangle}{\|f_1\|^2} f_1 = (-5, 4, 3, 2) - \frac{4}{30}(1, 2, 3, -4) = \frac{1}{15}(-77, 56, 39, 38),$$

where

$$\|f_2\|^2 = \frac{1}{225}((-77)^2 + 56^2 + 39^2 + 38^2) = \frac{12030}{225}.$$

Then an orthonormal basis of U is given by

$$e_1 = \frac{1}{\sqrt{30}}(1, 2, 3, -4), \quad e_2 = \frac{1}{\sqrt{12030}}(-77, 56, 39, 38).$$

An orthonormal basis for $U^\perp$ consists of vectors orthogonal to both v_1 and v_2. To find such vectors, consider the system of equations

$$\begin{cases} x + 2y + 3z - 4w = 0 \\ -5x + 4y + 3z + 2w = 0 \end{cases}$$

The first equation yields $x = -2y - 3z + 4w$, and substituting into the second equation gives

$$-5(-2y-3z+4w)+4y+3z+2w = 0 \implies 14y+18z-18w = 0 \implies 7y = -9z+9w.$$

Choosing $z = w = 1$ gives $y = 0$, and then $x = -2(0) - 3(1) + 4(1) = 1$. Choosing $z = 0$ and $w = 7$ gives $y = 9$ and then $x = -2(9) - 3(0) + 4(7) = -18 + 28 = 10$. Hence, we have two vectors $u_1 = (1, 0, 1, 1)$ and $u_2 = (10, 9, 0, 7)$ in $U^\perp$. Applying the Gram–Schmidt procedure to u_1 and u_2, we first set $g_1 = u_1$ with $\|g_1\|^2 = 3$. Then

$$g_2 = u_2 - \frac{\langle u_2, g_1 \rangle}{\|g_1\|^2} g_1 = (10, 9, 0, 7) - \frac{17}{3}(1, 0, 1, 1) = \frac{1}{3}(13, 27, -17, 4),$$

where

$$\|g_2\|^2 = \frac{1}{9}(13^2 + 27^2 + (-17)^2 + 4^2) = \frac{1203}{9}.$$

Then an orthonormal basis of $U^\perp$ is given by

$$e_1' = \frac{1}{\sqrt{3}}(1, 0, 1, 1), \quad e_2' = \frac{1}{\sqrt{1203}}(13, 27, -17, 4). \qquad \square$$

Solution 6C.4 Setting $v = e_j$, we have

$$1 = \|e_j\|^2 = |\langle e_j, e_1\rangle|^2 + \cdots + |\langle e_j, e_n\rangle|^2 = |\langle e_j, e_j\rangle|^2 + \sum_{k\neq j} |\langle e_j, e_k\rangle|^2.$$

Since $\|e_j\| = 1$, we have $|\langle e_j, e_j\rangle|^2 = 1$, so

$$1 = 1 + \sum_{k\neq j} |\langle e_j, e_k\rangle|^2 \implies \sum_{k\neq j} |\langle e_j, e_k\rangle|^2 = 0,$$

which forces $|\langle e_j, e_k\rangle|^2 = 0$ for each $k \neq j$, so $\langle e_j, e_k\rangle = 0$ for each $j \neq k$. Hence, $e_1, \ldots, e_n$ is an orthonormal list. To show that $e_1, \ldots, e_n$ is a basis of V, let $v \in V$ be arbitrary. Consider

$$u = v - \sum_{k=1}^{n} \langle v, e_k\rangle e_k.$$

Then for each j:

$$\langle u, e_j\rangle = \langle v, e_j\rangle - \sum_{k=1}^{n} \langle v, e_k\rangle\langle e_k, e_j\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0$$

by orthonormality. Thus, u is orthogonal to each e_j, so by the given condition,

$$\|u\|^2 = |\langle u, e_1\rangle|^2 + \cdots + |\langle u, e_n\rangle|^2 = 0,$$

which implies $u = 0$. Therefore, $v \in \operatorname{span}(e_1, \ldots, e_n)$. Since $e_1, \ldots, e_n$ is an orthonormal list and $\operatorname{span}(e_1, \ldots, e_n) = V$, it follows that $e_1, \ldots, e_n$ is an orthonormal basis of V. □

Solution 6C.5 For every $v \in V$, $v = P_U v + P_{U^\perp} v$ by the definition of orthogonal projection. Then

$$P_{U^\perp} v = v - P_U v = I v - P_U v = (I - P_U)v.$$

Hence, $P_{U^\perp} = I - P_U$. □

Solution 6C.6 We first show that $T = TP_{(\ker T)^\perp}$. For any $v \in V$, write $v = u + w$ for some $u \in \ker T$ and $w \in (\ker T)^\perp$. Then

$$TP_{(\ker T)^\perp} v = TP_{(\ker T)^\perp}(u + w) = TP_{(\ker T)^\perp} u + TP_{(\ker T)^\perp} w = T0 + Tw = Tw.$$

Since $u \in \ker T$, we have $Tu = 0$, so $Tv = T(u + w) = Tu + Tw = 0 + Tw = Tw$. Hence, $T = TP_{(\ker T)^\perp}$.
Next, we show that $T = P_{\operatorname{im} T} T$. For any $v \in V$, we have $Tv \in \operatorname{im} T$, so $P_{\operatorname{im} T} Tv = Tv$. Therefore, $T = P_{\operatorname{im} T} T$. □

Solution 6C.7 $\langle x, y\rangle = 0$ for all $x \in X, y \in Y \iff Y \subseteq X^\perp \iff \operatorname{im} P_Y \subseteq \ker P_X \iff P_X P_Y = 0$. □

Solution 6C.8 By the Riesz representation theorem, we have $w \in U$. For any $u \in U$, we have

$$\langle u, v - w \rangle = \langle u, v \rangle - \langle u, w \rangle = \phi(u) - \phi(u) = 0.$$

Hence, $v - w \in U^\perp$, so $w = P_U v$. □

Solution 6C.9 Let $U = \operatorname{im} P$. By Exercise 3B27, $P^2 = P$ implies $V = U \oplus \ker P$. Since every $v \in \ker P$ is orthogonal to every $u \in U$, we have $\ker P \subseteq U^\perp$. From the fundamental theorem of linear maps,

$$\dim \ker P = \dim V - \dim U = \dim U^\perp.$$

Hence, $\ker P = U^\perp$, so $V = U \oplus U^\perp$. Then for any $x \in V$, we can write $x = u + w$ for some $u \in U$ and $w \in U^\perp$. Since $u \in \operatorname{im} P$ and $P^2 = P$, we have $Pu = u$. Then $Px = P(u + w) = Pu + Pw = u + 0 = u = P_U x$, so $P = P_U$. □

Solution 6C.10 Let $U = \operatorname{im} P$. From Exercise 3B27, $P^2 = P$ implies $V = U \oplus \ker P$. Let $w \in \ker P$ and $u \in U$. For every $\alpha \in \mathbb{F}$,

$$\|u\| = \|P(u + \alpha w)\| \leq \|u + \alpha w\|.$$

By Exercise 6A6, $\langle u, w \rangle = 0$. Therefore, $\ker P \subseteq U^\perp$. From the fundamental theorem of linear maps,

$$\dim \ker P = \dim V - \dim U = \dim U^\perp.$$

Hence, $\ker P = U^\perp$, so $V = U \oplus U^\perp$. For any $v \in V$, write $v = u + w$ with $u \in U$ and $w \in U^\perp = \ker P$. Then $Pv = Pu + Pw = Pu = u = P_U v$, where $Pu = u$ since $u \in \operatorname{im} P$ and $P^2 = P$. Hence, $P = P_U$. □

Solution 6C.11 Suppose U is invariant under T. For any $v \in V$, we have $P_U v \in U$, so $TP_U v \in U$ by invariance. Since P_U acts as the identity on U, it follows that $P_U T P_U v = T P_U v$.
Conversely, suppose $P_U T P_U = T P_U$. For any $u \in U$, we have $P_U u = u$, so

$$Tu = TP_U u = P_U T P_U u \in \operatorname{im} P_U = U,$$

so U is invariant under T. □

Solution 6C.12 Suppose U and $U^\perp$ are both invariant under T. For any $v \in V$, write $v = u + w$ with $u \in U$ and $w \in U^\perp$. Then

$$P_U T v = P_U(Tu + Tw) = P_U Tu + P_U Tw = Tu + 0 = Tu,$$

since $Tu \in U$ and $Tw \in U^\perp$. On the other hand,

$$TP_U v = TP_U(u + w) = Tu,$$

so $P_U T v = TP_U v$ for all $v \in V$, and hence $P_U T = TP_U$.
Conversely, suppose $P_U T = TP_U$. For any $u \in U$, we have $P_U(Tu) = T(P_U u) = Tu$, so $Tu \in U$, and U is invariant under T. For any $w \in U^\perp$, we have $P_U(Tw) = T(P_U w) = T(0) = 0$, so $Tw \in U^\perp$, and $U^\perp$ is invariant under T. □

Solution 6C.13

(a) Define the map $\Phi : V \to V'$ by $\Phi(v) = \phi_v$ for all $v \in V$. To show that Φ is linear, let $v_1, v_2 \in V$ and $a, b \in \mathbb{R}$. Then for any $u \in V$,

$$\begin{aligned}\Phi(av_1 + bv_2)(u) &= \phi_{av_1+bv_2}(u)\\ &= \langle u, av_1 + bv_2\rangle\\ &= a\langle u, v_1\rangle + b\langle u, v_2\rangle\\ &= a\phi_{v_1}(u) + b\phi_{v_2}(u)\\ &= a\Phi(v_1)(u) + b\Phi(v_2)(u).\end{aligned}$$

Since this holds for all $u \in V$, we have $\Phi(av_1 + bv_2) = a\Phi(v_1) + b\Phi(v_2)$, so Φ is linear. Injectivity follows from the fact that if $\Phi(v) = 0$, then $\phi_v(u) = 0$ for all $u \in V$, which implies $\langle u, v\rangle = 0$ for all $u \in V$. In particular, taking $u = v$ gives $\langle v, v\rangle = 0$, so $v = 0$.

(b) By part (a), Φ is an injective linear map from V to V'. Since V is finite-dimensional, we have $\dim V = \dim V'$. Hence, Φ is also surjective, and therefore an isomorphism from V onto V'.

□

Solution 6C.14 Let $\phi_1, \ldots, \phi_n$ be the dual basis of $e_1, \ldots, e_n$. By the Riesz representation theorem, for each j there exists a unique $v_j \in V$ such that $\phi_j(u) = \langle u, v_j\rangle$ for all $u \in V$. Define $\psi_j \in V'$ by $\psi_j(u) = \langle u, e_j\rangle$. Then for any k,

$$\psi_j(e_k) = \langle e_k, e_j\rangle = \delta_{jk} = \phi_j(e_k).$$

Since ψ_j and ϕ_j agree on a basis of V, they are equal. By uniqueness in the Riesz representation theorem, $v_j = e_j$. Hence, the dual basis $\phi_1, \ldots, \phi_n$ corresponds to $e_1, \ldots, e_n$ under the identification of V' with V provided by the Riesz representation theorem. □

Solution 6C.15 Applying the Gram–Schmidt procedure to the basis $v_1 = (1, 1, 0, 0)$ and $v_2 = (1, 1, 1, 2)$ of U, we first set $f_1 = v_1$ with $\|f_1\|^2 = 2$. Then

$$f_2 = v_2 - \frac{\langle v_2, f_1\rangle}{\|f_1\|^2} f_1 = (1, 1, 1, 2) - \frac{2}{2}(1, 1, 0, 0) = (0, 0, 1, 2),$$

with $\|f_2\|^2 = 5$. Projecting $(1, 2, 3, 4)$ onto U gives

$$P_U(1, 2, 3, 4) = \frac{\langle (1, 2, 3, 4), f_1\rangle}{\|f_1\|^2} f_1 + \frac{\langle (1, 2, 3, 4), f_2\rangle}{\|f_2\|^2} f_2 = \frac{3}{2}(1, 1, 0, 0) + \frac{11}{5}(0, 0, 1, 2) = \left(\frac{3}{2}, \frac{3}{2}, \frac{11}{5}, \frac{22}{5}\right).$$

□

Solution 6C.16

Appendix. *An alternate proof is given at the end of this solution.*

(a) Let $g \in U^\perp$, and let $f(x) = x^2 g(x)$. Then f is continuous on $[-1,1]$ and $f(0) = 0$, so $f \in U$. Hence,

$$0 = \langle f, g \rangle = \int_{-1}^{1} x^2 g(x)^2 \, dx.$$

Since $x^2 g(x)^2 \geq 0$ for all $x \in [-1,1]$, it follows that $x^2 g(x)^2 = 0$ for all $x \in [-1,1]$. Therefore, $g(x) = 0$ for all $x \in [-1,1] \setminus \{0\}$, and by continuity, $g(0) = 0$ as well. Hence, $g = 0$, so $U^\perp = \{0\}$.

(b) 6.49 asserts that $V = U \oplus U^\perp$, but $U^\perp = \{0\}$, so $U \oplus U^\perp = U \neq C[-1,1]$ since the constant function 1 is not in U. 6.52 asserts that $(U^\perp)^\perp = U$, but $(U^\perp)^\perp = \{0\}^\perp = C[-1,1] \neq U$.

Alternate proof. A more analysis-based solution to part (a) is as follows.

Alternate proof. Let $g \in U^\perp$, so $\langle f, g \rangle = 0$ for all $f \in U$. For any $\varepsilon > 0$, define $f_\varepsilon \in C[-1,1]$ by $f_\varepsilon(x) = g(x)$ for $|x| \geq \varepsilon$, $f_\varepsilon(0) = 0$, with f_ε linear on $[-\varepsilon, 0]$ and on $[0, \varepsilon]$ so that f_ε is continuous at $x = 0$ and $x = \pm\varepsilon$. Then $f_\varepsilon \in U$, so $\langle f_\varepsilon, g \rangle = 0$. Let $M = \max_{[-1,1]} |g|$. Since $f_\varepsilon = g$ outside $[-\varepsilon, \varepsilon]$,

$$|\langle g, g \rangle| = |\langle f_\varepsilon - g, g \rangle| = \left| \int_{-\varepsilon}^{\varepsilon} (f_\varepsilon - g) g \right| \leq \int_{-\varepsilon}^{\varepsilon} |f_\varepsilon - g| \, |g| \leq 2M \cdot M \cdot 2\varepsilon = 4\varepsilon M^2.$$

Since $\varepsilon > 0$ is arbitrary, $\langle g, g \rangle = 0$, so $g = 0$. Hence, $U^\perp = \{0\}$. □

Solution 6C.17 The conditions $p(0) = 0$ and $p'(0) = 0$ imply that $p(x) = ax^2 + bx^3$ for some scalars $a, b \in \mathbb{R}$. Let $f(x) = 2 + 3x$ and $g(x) = ax^2 + bx^3$. To minimize the integral, we project f onto $\operatorname{span}(x^2, x^3)$ in $C[0,1]$ to get the resulting system of equations:

$$0 = \langle f - g, x^2 \rangle = \int_0^1 (2 + 3x - ax^2 - bx^3) x^2 \, dx = \frac{2}{3} + \frac{3}{4} - \frac{a}{5} - \frac{b}{6},$$
$$0 = \langle f - g, x^3 \rangle = \int_0^1 (2 + 3x - ax^2 - bx^3) x^3 \, dx = \frac{1}{2} + \frac{3}{5} - \frac{a}{6} - \frac{b}{7}.$$

Solving this system yields $a = 24$ and $b = -\frac{203}{10}$, so the polynomial $p(x) = 24x^2 - \frac{203}{10}x^3$ minimizes the integral. □

Solution 6C.18

Axler Note. *The polynomial 6.65 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 should help.*

Similar to Exercise 6C17, we project $\sin x$ onto $\operatorname{span}(1, x, x^2, x^3, x^4, x^5)$ in $C[-\pi, \pi]$. Note that $\sin x$ is orthogonal to 1, x^2, and x^4 on $[-\pi, \pi]$ because it is an odd function. Hence, we only need to consider the projection onto $\operatorname{span}(x, x^3, x^5)$. Using the Gram–Schmidt procedure, we can find an orthonormal basis of $\operatorname{span}(x, x^3, x^5)$ and then compute the projection of $\sin x$ onto this subspace. Calculating the orthogonal basis, we begin with $f_1 = x$ with $\|f_1\|^2 = \frac{2\pi^3}{3}$. Then

$$f_2 = x^3 - \frac{\langle x^3, f_1 \rangle}{\|f_1\|^2} f_1 = x^3 - \frac{3\pi^2}{5} x, \quad \|f_2\|^2 = \frac{8\pi^7}{175}.$$

Finally,

$$f_3 = x^5 - \frac{\langle x^5, f_1\rangle}{\|f_1\|^2} f_1 - \frac{\langle x^5, f_2\rangle}{\|f_2\|^2} f_2 = x^5 - \frac{10\pi^2}{9}x^3 + \frac{5\pi^4}{21}x, \quad \|f_3\|^2 = \frac{128\pi^{11}}{43659}.$$

Projecting $\sin x$ onto $U = \operatorname{span}(x, x^3, x^5)$ gives

$$P_U \sin x = \frac{\langle \sin x, f_1\rangle}{\|f_1\|^2} f_1 + \frac{\langle \sin x, f_2\rangle}{\|f_2\|^2} f_2 + \frac{\langle \sin x, f_3\rangle}{\|f_3\|^2} f_3,$$

where the inner products are computed as follows:

$$\begin{aligned}
\langle \sin x, f_1\rangle &= \int_{-\pi}^{\pi} x \sin x \, dx = 2\pi, \\
\langle \sin x, f_2\rangle &= \int_{-\pi}^{\pi} \left(x^3 - \frac{3\pi^2}{5}x\right) \sin x \, dx = \frac{4\pi}{5}(\pi^2 - 15), \\
\langle \sin x, f_3\rangle &= \int_{-\pi}^{\pi} \left(x^5 - \frac{10\pi^2}{9}x^3 + \frac{5\pi^4}{21}x\right) \sin x \, dx = \frac{2}{63}(8\pi^5 - 840\pi^3 + 7560\pi).
\end{aligned}$$

The coefficients of f_1, f_2, and f_3 in the projection are then

$$\frac{\langle \sin x, f_1\rangle}{\|f_1\|^2} = \frac{3}{\pi^2}, \quad \frac{\langle \sin x, f_2\rangle}{\|f_2\|^2} = \frac{35(\pi^2 - 15)}{2\pi^6}, \quad \frac{\langle \sin x, f_3\rangle}{\|f_3\|^2} = \frac{693(\pi^4 - 105\pi^2 + 945)}{8\pi^{10}}.$$

Hence, the polynomial $p(x) = P_U \sin x$ that minimizes the integral is $p(x) = \alpha x + \beta x^3 + \gamma x^5$, where

$$\begin{aligned}
\alpha &= \frac{105}{8\pi^6}(\pi^4 - 153\pi^2 + 1485), \\
\beta &= -\frac{315}{4\pi^8}(\pi^4 - 125\pi^2 + 1155), \\
\gamma &= \frac{693}{8\pi^{10}}(\pi^4 - 105\pi^2 + 945).
\end{aligned}$$

□

Solution 6C.19 For $u, v \in V$, write $u = u_U + u_{U^\perp}$ and $v = v_U + v_{U^\perp}$ with $u_U, v_U \in U = \operatorname{im} P$ and $u_{U^\perp}, v_{U^\perp} \in U^\perp = \ker P$. Then $Pu = u_U$ and $Pv = v_U$, so

$$\langle Pu, v\rangle = \langle u_U, v_U + v_{U^\perp}\rangle = \langle u_U, v_U\rangle$$

since $u_U \in U$ and $v_{U^\perp} \in U^\perp$. On the other hand,

$$\langle u, Pv\rangle = \langle u_U + u_{U^\perp}, v_U\rangle = \langle u_U, v_U\rangle$$

since $u_{U^\perp} \in U^\perp$ and $v_U \in U$. Hence, $\langle Pu, v\rangle = \langle u, Pv\rangle$ for all $u, v \in V$, so $P^\dagger = P$. □

Solution 6C.20 Let P be the orthogonal projection of W onto $\operatorname{im} T$ and let $\tilde{T} : (\ker T)^\perp \to \operatorname{im} T$ denote the restriction of T to $(\ker T)^\perp$, so that $T^\dagger = \tilde{T}^{-1}P$. Since $\tilde{T}^{-1} : \operatorname{im} T \to (\ker T)^\perp$ is a bijection, and P maps W onto $\operatorname{im} T$,

$$\ker T^\dagger = \ker(\tilde{T}^{-1}P) = \ker P = (\operatorname{im} T)^\perp,$$

where the middle equality uses injectivity of $\tilde{T}^{-1}$, and

$$\operatorname{im} T^\dagger = \tilde{T}^{-1}(P(W)) = \tilde{T}^{-1}(\operatorname{im} T) = (\ker T)^\perp,$$

where the second equality uses surjectivity of P onto $\operatorname{im} T$. □

Solution 6C.21

(a) From $a + b + c = 2b + 3c = 0$, we see that $\ker T = \operatorname{span}(1, -3, 2)$, so

$$(\ker T)^\perp = \{(a, b, c) : a - 3b + 2c = 0\}.$$

Moreover, $\operatorname{im} T = \mathbb{F}^2$, so $P_{\operatorname{im} T} = I$. Let $T^\dagger(x, y) = (a, b, c)$. Then $T(a, b, c) = (x, y)$ and $(a, b, c) \in (\ker T)^\perp$. Hence, we have the system of equations

$$\begin{cases} x = a + b + c, \\ y = 2b + 3c, \\ 0 = a - 3b + 2c. \end{cases}$$

This yields the solution

$$T^\dagger(x, y) = \left(\frac{13x - 5y}{14}, \frac{3x + y}{14}, \frac{-2x + 4y}{14}\right).$$

(b) For $(x, y) \in \mathbb{F}^2$, we have

$$T T^\dagger(x, y) = T\left(\frac{13x - 5y}{14}, \frac{3x + y}{14}, \frac{-2x + 4y}{14}\right) = (x, y),$$

as required.

(c) We have $P_{(\ker T)^\perp}$ being the projection onto the plane $a - 3b + 2c = 0$. Hence,

$$P_{(\ker T)^\perp}(a, b, c) = (a, b, c) - \frac{a - 3b + 2c}{14}(1, -3, 2) = \left(\frac{13a + 3b - 2c}{14}, \frac{3a + 5b + 6c}{14}, \frac{-2a + 6b + 10c}{14}\right).$$

For $(a, b, c) \in \mathbb{F}^3$, we have

$$T^\dagger T(a, b, c) = T^\dagger(a+b+c, 2b+3c) = \left(\frac{13a + 3b - 2c}{14}, \frac{3a + 5b + 6c}{14}, \frac{-2a + 6b + 10c}{14}\right),$$

as required.

□

Solution 6C.22 Let $V = \ker T \oplus (\ker T)^\perp$ so that any $v \in V$ can be written as $v = u + w$ with $u \in \ker T$ and $w \in (\ker T)^\perp$. For the first equality, we have $Tv = Tw \in \operatorname{im} T$. Then $T^\dagger T w = w$ since $T^\dagger$ restricted to $\operatorname{im} T$ is the inverse of T restricted to $(\ker T)^\perp$. Hence,

$$T T^\dagger T v = T T^\dagger T w = T w = T v$$

for all $v \in V$, so $T T^\dagger T = T$.

For the second equality, let $w \in W$ and write $w = x + y$ with $x \in \operatorname{im} T$ and $y \in (\operatorname{im} T)^\perp$. Then $T^\dagger w = T^\dagger x \in (\ker T)^\perp$, so

$$T^\dagger T T^\dagger w = T^\dagger T(T^\dagger x) = T^\dagger x = T^\dagger w.$$

Hence, $T^\dagger T T^\dagger = T^\dagger$. □

Solution 6C.23 By Exercise 6C20, $\ker T^\dagger = (\operatorname{im} T)^\perp$ and $\operatorname{im} T^\dagger = (\ker T)^\perp$. Applying 6.68 to $T^\dagger$ in place of T and substituting via Exercise 6C20, for $v \in V$,

$$(T^\dagger)^\dagger v = (T^\dagger|_{(\ker T^\dagger)^\perp})^{-1} P_{\operatorname{im} T^\dagger} v = (T^\dagger|_{\operatorname{im} T})^{-1} P_{(\ker T)^\perp} v.$$

By 6.67, $T|_{(\ker T)^\perp}$ is an isomorphism onto $\operatorname{im} T$ whose inverse is $T^\dagger|_{\operatorname{im} T}$, so $(T^\dagger|_{\operatorname{im} T})^{-1} = T|_{(\ker T)^\perp}$. Because $v - P_{(\ker T)^\perp} v \in ((\ker T)^\perp)^\perp = \ker T$, we have $T(v - P_{(\ker T)^\perp} v) = 0$, so

$$(T^\dagger)^\dagger v = T(P_{(\ker T)^\perp} v) = Tv. \qquad \square$$

Chapter 7

§7A

Solution 7A.1 If $n = 1$, then $T(z_1) = 0$, so $T^*(z_1) = 0$. Let $n > 1$. Then for all $(z_1, \ldots, z_n), (w_1, \ldots, w_n) \in \mathbb{F}^n$, we have

$$\langle T(z_1, \ldots, z_n), (w_1, \ldots, w_n)\rangle = \langle (0, z_1, \ldots, z_{n-1}), (w_1, \ldots, w_n)\rangle = z_1\overline{w_2} + \cdots + z_{n-1}\overline{w_n}.$$

We need

$$\langle (z_1, \ldots, z_n), T^*(w_1, \ldots, w_n)\rangle = z_1\overline{w_2} + \cdots + z_{n-1}\overline{w_n}.$$

By comparison, we see that $T^*(w_1, \ldots, w_n) = (w_2, \ldots, w_n, 0)$, so $T^*(z_1, \ldots, z_n) = (z_2, \ldots, z_n, 0)$ for all $(z_1, \ldots, z_n) \in \mathbb{F}^n$. $\square$

Solution 7A.2 Since $(T^*)^* = T$, then $T = 0 \iff T^* = 0$. Moreover, if $T^* = 0$, then $T^*T = 0$. Conversely, if $T^*T = 0$, then for all $v \in V$, we have

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle v, T^*Tv\rangle = \langle v, 0\rangle = 0.$$

Hence, $Tv = 0$ for all $v \in V$, so $T = 0$. Therefore,

$$T = 0 \iff T^* = 0 \iff T^*T = 0.$$

Replacing T with T^* in the equivalence $T = 0 \iff T^*T = 0$ gives $T^* = 0 \iff TT^* = 0$. We then have

$$T = 0 \iff T^* = 0 \iff T^*T = 0 \iff TT^* = 0. \qquad \square$$

Solution 7A.3 Since V is finite-dimensional,

$$\dim \ker(T - \lambda I) = \dim V - \dim \operatorname{im}(T - \lambda I) = \dim(\operatorname{im}(T - \lambda I)^\perp) = \dim \ker((T - \lambda I)^*).$$

Hence,

$$\ker(T - \lambda I) \neq \{0\} \iff \ker((T - \lambda I)^*) \neq \{0\}.$$

Since $(T - \lambda I)^* = T^* - \overline{\lambda} I$, we have

$$\lambda \text{ is an eigenvalue of } T \iff \overline{\lambda} \text{ is an eigenvalue of } T^*. \qquad \square$$

Solution 7A.4

$$\begin{aligned}
U^\perp \text{ is invariant under } T^* &\iff T^*v \in U^\perp \text{ for all } v \in U^\perp \\
&\iff \langle u, T^*v\rangle = 0 \text{ for all } u \in U, v \in U^\perp \\
&\iff \langle Tu, v\rangle = 0 \text{ for all } u \in U, v \in U^\perp \\
&\iff Tu \in (U^\perp)^\perp \text{ for all } u \in U \\
&\iff Tu \in U \text{ for all } u \in U \\
&\iff U \text{ is invariant under } T
\end{aligned}$$

□

Solution 7A.5

Axler Note. *The numbers $\|Te_1\|^2, \ldots, \|Te_n\|^2$ in the equation above depend on the orthonormal basis $e_1, \ldots, e_n$, but the right side of the equation does not depend on $e_1, \ldots, e_n$. Thus, the equation above shows that the sum on the left side does not depend on which orthonormal basis $e_1, \ldots, e_n$ is used.*

Manual Remark. *For students familiar with this topic already, note that the left-hand side is the squared Frobenius norm of the matrix representation of T with respect to the orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$. This exercise shows that $\sum_i \|Te_i\|^2 = \sum_j \|T^* f_j\|^2$, which is independent of the choice of orthonormal basis $e_1, \ldots, e_n$. Moreover, noting that $\operatorname{tr}(T^*T) = \operatorname{tr}(TT^*)$ by the cyclic property of the trace, we see that the sum of the squared norms of the columns of the matrix representation of T equals the sum of the squared norms of the rows, which aligns with the result of this exercise.*

Let

$$Te_j = \sum_{k=1}^{m} \langle Te_j, f_k\rangle f_k \implies \|Te_j\|^2 = \sum_{k=1}^{m} |\langle Te_j, f_k\rangle|^2.$$

Similarly, we may write

$$T^* f_i = \sum_{\ell=1}^{n} \langle T^* f_i, e_\ell\rangle e_\ell \implies \|T^* f_i\|^2 = \sum_{\ell=1}^{n} |\langle T^* f_i, e_\ell\rangle|^2.$$

Note that

$$\langle T^* f_i, e_j\rangle = \langle f_i, Te_j\rangle = \overline{\langle Te_j, f_i\rangle} \implies |\langle T^* f_i, e_j\rangle|^2 = |\langle Te_j, f_i\rangle|^2.$$

Therefore, we have

$$\sum_{k=1}^{m} \|T^* f_k\|^2 = \sum_{k=1}^{m}\sum_{\ell=1}^{n} |\langle Te_\ell, f_k\rangle|^2 = \sum_{\ell=1}^{n}\sum_{k=1}^{m} |\langle Te_\ell, f_k\rangle|^2 = \sum_{\ell=1}^{n} \|Te_\ell\|^2. \qquad \square$$

Solution 7A.6

(a) T injective $\iff \ker T = \{0\} \iff (\operatorname{im} T^*)^\perp = \{0\} \iff \operatorname{im} T^* = V \iff T^*$ surjective.

(b) Using part (a) on $T^* \in \mathrm{Hom}(W, V)$ and the fact that $(T^*)^* = T$, we have

$$T^* \text{ is injective} \iff (T^*)^* \text{ is surjective} \iff T \text{ is surjective.}$$

□

Solution 7A.7

(a) By Theorem 7.6(a), we have $\ker T^* = (\operatorname{im} T)^\perp$, so $\dim \ker T^* = \dim W - \dim \operatorname{im} T$. Using the fundamental theorem of linear maps, we have $\dim \operatorname{im} T = \dim V - \dim \ker T$, so

$$\dim \ker T^* = \dim W - \dim V + \dim \ker T.$$

(b) We have $\operatorname{im} T^* = (\ker T)^\perp$, so $\dim \operatorname{im} T^* = \dim V - \dim \ker T = \dim \operatorname{im} T$.

□

Solution 7A.8 Let $T \in \mathrm{Hom}(\mathbb{F}^n, \mathbb{F}^m)$ have the matrix $A = \mathrm{Mat}(T)$ with respect to the standard bases of $\mathbb{F}^n$ and $\mathbb{F}^m$. Since the standard bases of $\mathbb{F}^n$ and $\mathbb{F}^m$ are orthonormal, $A^* = \mathrm{Mat}(T^*)$ are the conjugate-transposed rows of A. Since complex conjugation preserves linear independence, the column rank of A^* equals the row rank of A. By Exercise 7A7(b), $\dim \operatorname{im} T^* = \dim \operatorname{im} T$, so the column rank of A^* equals the column rank of A, and the result follows. □

Solution 7A.9 Let $S, T \in \mathrm{End}(V)$ be self-adjoint, so $S^* = S$ and $T^* = T$. Note that $(ST)^* = T^*S^* = TS$. Hence,

$$ST \text{ is self-adjoint} \iff (ST)^* = ST \iff TS = ST. \qquad \square$$

Solution 7A.10 $\Rightarrow$ Suppose $\langle Tv, v\rangle = \langle T^*v, v\rangle$ for all $v \in V$. Then for all $v \in V$,

$$\langle (T - T^*)v, v\rangle = 0.$$

Since $\mathbb{F} = \mathbb{C}$, we may use 7.13 with $T - T^*$ in place of T to conclude that $T - T^* = 0$, so $T = T^*$, i.e., T is self-adjoint.
$\Leftarrow$ If T is self-adjoint, then $T = T^*$, so for all $v \in V$,

$$\langle Tv, v\rangle = \langle T^*v, v\rangle. \qquad \square$$

Solution 7A.11

Axler Note. *If $\mathbb{F} = \mathbb{R}$, then S is the operator on $\mathbb{R}^2$ of counterclockwise rotation by* 90°.

(a) For any $(w,z),(x,y) \in \mathbb{F}^2$,

$$\langle S(w,z),(x,y)\rangle = \langle(-z,w),(x,y)\rangle = -z\overline{x} + w\overline{y}.$$

Suppose $S^*(x,y) = (u,v)$. Then

$$\langle S(w,z),(x,y)\rangle = \langle(w,z),S^*(x,y)\rangle = \langle(w,z),(u,v)\rangle = w\overline{u} + z\overline{v}.$$

Comparing with $-z\overline{x} + w\overline{y}$, we get $w\overline{u} + z\overline{v} = -z\overline{x} + w\overline{y}$ for all $w,z \in \mathbb{F}$. Hence, $u = y$ and $v = -x$, so $S^*(x,y) = (y,-x)$.

(b) Observe that

$$SS^*(w,z) = S(S^*(w,z)) = S(z,-w) = (w,z),$$

and

$$S^*S(w,z) = S^*(S(w,z)) = S^*(-z,w) = (w,z).$$

Hence, $SS^* = S^*S$, so S is normal. However,

$$S(1,0) = (0,1) \neq (0,-1) = S^*(1,0).$$

Hence, S is not self-adjoint.

(c) If $S(w,z) = \lambda(w,z)$ for some $\lambda \in \mathbb{F}$ and $(w,z) \neq (0,0)$, then $-z = \lambda w$ and $w = \lambda z$. From this, we get $w = -\lambda^2 w$, so $\lambda^2 = -1$. If $\mathbb{F} = \mathbb{R}$, then S has no real eigenvalues. If $\mathbb{F} = \mathbb{C}$, then the eigenvalues of S are $\lambda = i$ and $\lambda = -i$.

□

Solution 7A.12 $\Rightarrow$ Suppose there exist commuting operators A and B such that A is self-adjoint, B is skew, and $T = A + B$. Then

$$T^* = (A+B)^* = A^* + B^* = A - B.$$

Hence,

$$TT^* = (A+B)(A-B) = A^2 - AB + BA - B^2 = A^2 - B^2,$$

and

$$T^*T = (A-B)(A+B) = A^2 + AB - BA - B^2 = A^2 - B^2.$$

Therefore, $TT^* = T^*T$, so T is normal.
$\Leftarrow$ If T is normal, then $TT^* = T^*T$. Consider the operators

$$A = \frac{T+T^*}{2}, \quad B = \frac{T-T^*}{2}.$$

Then A is self-adjoint, B is skew, and $T = A + B$. Moreover,

$$\begin{aligned} AB &= \frac{T+T^*}{2} \cdot \frac{T-T^*}{2} \\ &= \frac{1}{4}(TT - TT^* + T^*T - T^*T^*) \\ &= \frac{1}{4}(TT + TT^* - T^*T - T^*T^*) \\ &= \frac{T-T^*}{2} \cdot \frac{T+T^*}{2} = BA, \end{aligned}$$

hence A and B commute. □

Solution 7A.13

(a) If $V = \{0\}$, then $\text{End}(V) = \{0\}$ and $\mathcal{A}$ has no eigenvalues. Otherwise, assume $\dim V \geq 1$. If $\mathcal{A}T = \lambda T$ for some nonzero $T \in \text{End}(V)$ and $\lambda \in \mathbb{R}$, then $T^* = \lambda T$. Taking the adjoint of both sides gives

$$T = \lambda T^* = \lambda(\lambda T) = \lambda^2 T.$$

Hence, $\lambda^2 = 1$, so $\lambda = 1$ or $\lambda = -1$. To see that $\lambda = 1$ always occurs as an eigenvalue for $\dim V \geq 1$, note that $I^* = I$, so $\mathcal{A}I = I = 1 \cdot I$. If $\dim V = 1$, then $T = \mu I$ for some $\mu \in \mathbb{R}$. In this case, $T^* = T$, so $\mathcal{A}T = T = 1 \cdot T$, and the only eigenvalue of $\mathcal{A}$ is 1.

If $\dim V \geq 2$, let $e_1, \ldots, e_n$ be an orthonormal basis of V with $n = \dim V$. Define $S \in \text{End}(V)$ by $Se_1 = e_2$, $Se_2 = -e_1$, and $Se_j = 0$ for $j \geq 3$. Then $S^*e_1 = -e_2$, $S^*e_2 = e_1$, and $S^*e_j = 0$ for $j \geq 3$, so $S^* = -S$. Hence, -1 is also an eigenvalue of $\mathcal{A}$. Therefore, the eigenvalues of $\mathcal{A}$ are 1 and -1 when $\dim V \geq 2$.

(b) Observe that $\mathcal{A}(\mathcal{A}T) = \mathcal{A}T^* = (T^*)^* = T$ for all $T \in \text{End}(V)$. Hence, $\mathcal{A}^2 - I = 0$, so the minimal polynomial of $\mathcal{A}$ must divide $z^2 - 1$. If $V = \{0\}$, then $\text{End}(V) = \{0\}$, and the minimal polynomial of $\mathcal{A}$ is 1. If $\dim V = 1$, then $\mathcal{A}$ has only one eigenvalue, 1, so the minimal polynomial of $\mathcal{A}$ is $z - 1$. Otherwise, both eigenvalues 1 and -1 occur when $\dim V \geq 2$, so the minimal polynomial of $\mathcal{A}$ is $z^2 - 1$.

□

Solution 7A.14

(a) Consider $p = 1$ and $q = x$. Then

$$\langle Tp, q \rangle = \langle 0, x \rangle = 0,$$

but

$$\langle p, Tq \rangle = \langle 1, x \rangle = \int_0^1 x \, dx = \frac{1}{2} \neq 0.$$

Hence, T is not self-adjoint.

(b) Observe that

$$\|x\|^2 = \langle x, x \rangle = \int_0^1 x^2 \, dx = \frac{1}{3}.$$

By 7.9, given an orthonormal basis, T is self-adjoint if and only if the matrix of T with respect to that basis equals its conjugate transpose. However, the basis $1, x, x^2$ is not orthonormal with respect to the given inner product, so the fact that the matrix of T with respect to this basis equals its conjugate transpose does not imply that T is self-adjoint.

□

Solution 7A.15 We first note that

$$TT^{-1} = I \implies (T^{-1})^* T^* = I^* = I \implies (T^{-1})^* = (T^*)^{-1}.$$

(a)

$$\begin{aligned} T^{-1} \text{ is self-adjoint} &\iff (T^{-1})^* = T^{-1} \\ &\iff (T^*)^{-1} = T^{-1} \\ &\iff T^* = T \\ &\iff T \text{ is self-adjoint.} \end{aligned}$$

(b)

$$\begin{aligned} T^{-1} \text{ is normal} &\iff (T^{-1})^* T^{-1} = T^{-1}(T^{-1})^* \\ &\iff (T^*)^{-1} T^{-1} = T^{-1}(T^*)^{-1} \\ &\iff (T^* T)^{-1} = (T T^*)^{-1} \\ &\iff T^* T = T T^* \\ &\iff T \text{ is normal.} \end{aligned}$$

□

Solution 7A.16

(a) Let $\mathcal{S}$ be the set of self-adjoint operators on V. Since $0^* = 0$, we have $0 \in \mathcal{S}$. For $S, T \in \mathcal{S}$ and $\lambda \in \mathbb{R}$, we have

$$(S + T)^* = S^* + T^* = S + T \in \mathcal{S},$$

and

$$(\lambda T)^* = \overline{\lambda} T^* = \lambda T \in \mathcal{S}.$$

Therefore, $\mathcal{S}$ is closed under addition and scalar multiplication, and contains the zero operator, so it is a subspace of $\text{End}(V)$.

(b) Let $n = \dim V$, and let $e_1, \ldots, e_n$ be an orthonormal basis of V. By 7.9, it follows that T is self-adjoint if and only if $\text{Mat}(T)$ is a symmetric matrix, i.e., $\text{Mat}(T) = \text{Mat}(T)^{\text{t}}$. Since a symmetric $n \times n$ matrix is determined by the entries on and above the main diagonal, the dimension of the subspace of symmetric matrices is

$$\frac{n(n+1)}{2}.$$

□

Solution 7A.17 Note the result is not true when $V = \{0\}$, since the only operator on V is the zero operator, which is self-adjoint. Let $\mathcal{S}$ be the set of self-adjoint operators on V. If $\mathcal{S}$ were a subspace of $\text{End}(V)$, consider iI, where $I \in \mathcal{S}$ is the identity operator on V. Then

$$(iI)^* = \overline{i} I = -iI \neq iI.$$

Therefore, iI is not self-adjoint, which shows that the set of self-adjoint operators on V is not a subspace of $\text{End}(V)$ when $V \neq \{0\}$. □

Solution 7A.18 Fix an orthonormal pair $e_1, e_2 \in V$. Extend this to an orthonormal basis $e_1, \ldots, e_n$ of V. Consider the operators S and T defined by

$$Se_1 = e_2, Se_2 = e_1, Se_i = 0 \text{ for } i > 2, \quad Te_1 = -e_2, Te_2 = e_1, Te_i = 0 \text{ for } i > 2.$$

Then

$$\operatorname{Mat}(S) = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

From this, we see that $\operatorname{Mat}(S) = \operatorname{Mat}(S)^{\mathrm{t}}$, so S is normal. Moreover,

$$\operatorname{Mat}(T) = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Direct computation yields $\operatorname{Mat}(T)\operatorname{Mat}(T)^* = \operatorname{Mat}(T)^*\operatorname{Mat}(T)$, so T is normal. However, $S + T$ is not normal, as can be seen by computing

$$\operatorname{Mat}(S+T) = \begin{pmatrix} 0 & 2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \operatorname{Mat}(S+T)^* = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Direct computation shows that $\operatorname{Mat}(S+T)\operatorname{Mat}(S+T)^* \neq \operatorname{Mat}(S+T)^*\operatorname{Mat}(S+T)$, so $S + T$ is not normal. □

Solution 7A.19 If $\dim V = 0$, then T is trivially normal. So assume $\dim V > 0$. Note that for all $v \in V$, we have

$$\|T^*v\|^2 = \langle T^*v, T^*v\rangle = \langle TT^*v, v\rangle, \quad \|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle.$$

From the hypothesis, we have $\|T^*v\|^2 \leq \|Tv\|^2$ for all $v \in V$, which gives

$$\langle (T^*T - TT^*)v, v\rangle \geq 0 \quad \text{for all } v \in V.$$

Let $S = T^*T - TT^*$. Then $S^* = (T^*T - TT^*)^* = T^*T - TT^* = S$, so S is self-adjoint. Moreover, $\langle Sv, v\rangle \geq 0$ for all $v \in V$. Let $e_1, \ldots, e_n$ be an orthonormal basis of V. From Exercise 7A5, then

$$\sum_{i=1}^{n} \|Te_i\|^2 = \sum_{i=1}^{n} \langle T^*Te_i, e_i\rangle = \sum_{i=1}^{n} \langle TT^*e_i, e_i\rangle = \sum_{i=1}^{n} \|T^*e_i\|^2.$$

Then

$$\sum_{i=1}^{n} \langle Se_i, e_i\rangle = \sum_{i=1}^{n} \langle (T^*T - TT^*)e_i, e_i\rangle = \sum_{i=1}^{n} \langle T^*Te_i, e_i\rangle - \sum_{i=1}^{n} \langle TT^*e_i, e_i\rangle = 0.$$

Since each $\langle Se_i, e_i\rangle \geq 0$ and their sum is 0, it follows that $\langle Se_i, e_i\rangle = 0$ for all i. For arbitrary $u \in V$ with $\|u\| = 1$, we may extend it to an orthonormal basis $u, f_2, \ldots, f_n$ of V. Then

$$\langle Su, u\rangle + \sum_{i=2}^{n} \langle Sf_i, f_i\rangle = \sum_{i=1}^{n} \langle Se_i, e_i\rangle = 0.$$

Since each term is nonnegative, it follows that $\langle Su, u\rangle = 0$ for all unit vectors $u \in V$. By homogeneity, $\langle Sv, v\rangle = 0$ for all $v \in V$. From 7.16, if S is self-adjoint and $\langle Sv, v\rangle = 0$ for all $v \in V$, then $S = 0$. Hence, $TT^* - T^*T = 0$, so $TT^* = T^*T$, and therefore T is normal. □

Solution 7A.20
(a) $\implies$ (b): If P is self-adjoint, then $P = P^*$. Hence,

$$PP^* = PP = P^2 = P = P^*P,$$

so P is normal.
(b) $\implies$ (c): Since P is normal,

$$(\operatorname{im} P)^{\perp} = \ker P^* = \ker P.$$

Since $P^2 = P$, $V = \ker P \oplus \operatorname{im} P$ is an orthogonal direct sum. Letting $U = \operatorname{im} P$, then any $v \in V$ can be written uniquely as $v = u + w$ with $u \in U$ and $w \in \ker P$. Hence,

$$Pv = P(u + w) = Pu + Pw = u + 0 = u \in U.$$

Therefore, $P = P_U$.
(c) $\implies$ (a): Write $V = U \oplus U^{\perp}$. Then for $v = u_1 + w_1$ and $w = u_2 + w_2$ with $u_1, u_2 \in U$ and $w_1, w_2 \in U^{\perp}$, we have

$$\langle P_U v, w\rangle = \langle u_1, u_2 + w_2\rangle = \langle u_1, u_2\rangle = \langle u_1 + w_1, u_2\rangle = \langle v, P_U w\rangle.$$

Hence, P_U is self-adjoint. □

Solution 7A.21 If there did exist an inner product on $\mathbb{R}[X]_{\leq 8}$ that makes D a normal operator, then $\|Dp\| = \|D^*p\|$ for all $p \in \mathbb{R}[X]_{\leq 8}$, by 7.20. Consider $p = x$ so that $Dp = 1 \neq 0$, but $D^2p = 0$. Then applying 7.20 to Dp, we get

$$\|D^*(Dp)\| = \|D(Dp)\| = \|D^2p\| = \|0\| = 0$$

so that $D^*(Dp) = 0$. Then

$$\langle D^*(Dp), p\rangle = \langle Dp, Dp\rangle = \|Dp\|^2.$$

Since $D^*(Dp) = 0$, the left-hand side of the above equation is 0, while the right-hand side is $\|Dp\|^2 \neq 0$. This is a contradiction, so no such inner product can exist. □

Solution 7A.22 Let T have the matrix

$$\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with respect to the standard basis of $\mathbb{R}^3$, which is orthonormal. Then T is normal because

$$\mathrm{Mat}(T^*) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathrm{Mat}(T)^*.$$

By direct computation, we have

$$TT^* = I, \quad T^*T = I.$$

Hence, T is normal but not self-adjoint, since $T \neq T^*$. □

Solution 7A.23 Since v, w are eigenvectors of the normal operator T corresponding to distinct eigenvalues 3 and 4, they must be orthogonal. Therefore, $\langle v, w\rangle = 0$. Then

$$\begin{aligned} \|T(v+w)\|^2 &= \langle Tv + Tw, Tv + Tw\rangle \\ &= \langle 3v + 4w, 3v + 4w\rangle \\ &= \langle 3v, 3v\rangle + \langle 3v, 4w\rangle + \langle 4w, 3v\rangle + \langle 4w, 4w\rangle \\ &= 9\langle v, v\rangle + 12\langle v, w\rangle + 12\langle w, v\rangle + 16\langle w, w\rangle \\ &= 9\|v\|^2 + 16\|w\|^2 \\ &= 9 \cdot 4 + 16 \cdot 4 \\ &= 100. \end{aligned}$$

Therefore, $\|T(v+w)\| = 10$. □

Solution 7A.24 Define $p(z)$ to be the minimal polynomial of T. Let $q(z) = \sum_{i=0}^{m} \overline{a_i} z^i$. Then

$$q(T^*) = \sum_{i=0}^{m} \overline{a_i} (T^*)^i = \sum_{i=0}^{m} \overline{a_i} (T^i)^* = \left(\sum_{i=0}^{m} a_i T^i \right)^* = p(T)^* = 0.$$

Hence, $q(z)$ is a polynomial that annihilates T^*. Using a similar argument with $r(z)$ being the minimal polynomial of T^*, we can show that $\overline{r}(T) = 0$. Since $q(z)$ annihilates T^* and $r(z)$ is the minimal polynomial of T^*, it follows that $r(z)$ divides $q(z)$. Similarly, since $\overline{r}(T) = 0$, $p(z)$ divides $\overline{r}(z)$. Then $\deg r \leq \deg q$ from the fact that $r(z)$ divides $q(z)$, and $\deg q \leq \deg r$ from the fact that $p(z)$ divides $\overline{r}(z)$ and $\deg p = \deg q$. Hence, $\deg r = \deg q$, and since $r(z)$ is monic and divides $q(z)$, we must have $r(z) = q(z)$. Therefore, the minimal polynomial of T^* is $q(z)$. □

Solution 7A.25 $\Rightarrow$ Suppose T is diagonalizable. Then the minimal polynomial of T splits into distinct linear factors by 5.62. By Exercise 7A24, the minimal polynomial of T^* also splits into distinct linear factors whose roots are the complex conjugates of the roots of the minimal polynomial of T. Hence, T^* is diagonalizable.
$\Leftarrow$ Conversely, suppose T^* is diagonalizable. Apply the same argument as above, using the fact that $(T^*)^* = T$. Therefore, T is diagonalizable. □

Solution 7A.26 Note that

$$\langle Tv, w\rangle = \langle \langle v, u\rangle x, w\rangle = \langle v, u\rangle\langle x, w\rangle = \langle v, \overline{\langle x, w\rangle}u\rangle = \langle v, T^*w\rangle.$$

Since $\langle v, u_1\rangle = \langle v, u_2\rangle$ for all $v \in V$ implies $u_1 = u_2$, we conclude that $T^*w = \overline{\langle x, w\rangle}u = \langle w, x\rangle u$.

(a) $\Rightarrow$ If T is self-adjoint, then $T = T^*$, which means $\langle v, u\rangle x = \langle v, x\rangle u$ for all $v \in V$. If $u = 0$ or $x = 0$, then the list u, x is linearly dependent. If both are nonzero, then letting $v = u$ gives $\|u\|^2 x = \langle u, x\rangle u$, hence

$$x = \frac{\langle u, x\rangle}{\|u\|^2}u.$$

Hence, x and u are linearly dependent.

$\Leftarrow$ Conversely, suppose one of u or x is zero. Then $T = 0 = T^*$, so T is self-adjoint. If both are nonzero and linearly dependent, let $x = \lambda u$ for some $\lambda \in \mathbb{R}$. Then for all $v \in V$,

$$Tv = \langle v, u\rangle x = \langle v, u\rangle \lambda u = \lambda\langle v, u\rangle u = T^*v,$$

hence T is self-adjoint.

(b) $\Rightarrow$ Suppose T is normal. Then for any $y \in V$,

$$TT^*y = T(\langle y, x\rangle u) = \langle y, x\rangle Tu = \langle y, x\rangle\langle u, u\rangle x = \|u\|^2\langle y, x\rangle x,$$

and

$$T^*Ty = T^*(\langle y, u\rangle x) = \langle y, u\rangle T^*x = \langle y, u\rangle\langle x, x\rangle u = \|x\|^2\langle y, u\rangle u.$$

If $u = 0$ or $x = 0$, then the list u, x is linearly dependent. Otherwise, for nonzero u and x, we must have $\|u\|^2\langle y, x\rangle x = \|x\|^2\langle y, u\rangle u$ for all $y \in V$. Let $y = u$. Then $\|u\|^2\langle u, x\rangle x = \|x\|^2\|u\|^2 u$, hence $\langle u, x\rangle x = \|x\|^2 u$. If $x \neq 0$, then

$$u = \frac{\langle u, x\rangle}{\|x\|^2}x,$$

so u and x are linearly dependent.

$\Leftarrow$ If one of u or x is zero, then $T = 0 = T^*$, so T is normal. If both are nonzero and linearly dependent, let $u = \mu x$ for some $\mu \in \mathbb{F}$. Then for all $y \in V$,

$$TT^*y = \|u\|^2\langle y, x\rangle x = \|\mu x\|^2\langle y, x\rangle x = |\mu|^2\|x\|^2\langle y, x\rangle x,$$

and

$$T^*Ty = \|x\|^2\langle y, u\rangle u = \|x\|^2\langle y, \mu x\rangle \mu x = |\mu|^2\|x\|^2\langle y, x\rangle x.$$

Hence, $TT^*y = T^*Ty$ for all $y \in V$, so T is normal. □

Solution 7A.27 Let $v \in \ker T$. Then $T^k v = 0$, so $v \in \ker T^k$. Hence, $\ker T \subseteq \ker T^k$. Conversely, let $v \in \ker T^k$. Since T is normal, $\|Tv\| = \|T^*v\|$ by 7.20. Hence, $\ker T = \ker T^*$. If $T^2 v = 0$, then $T(Tv) = 0$, so $Tv \in \ker T = \ker T^*$. Therefore, $T^*Tv = 0$ so that $\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = 0$. Hence, $Tv = 0$, which shows that $v \in \ker T$. By induction on k, assume that $\ker T^{k-1} = \ker T$. If $v \in \ker T^k$, then $Tv \in \ker T^{k-1} = \ker T$, so $T(Tv) = T^2 v = 0$. By the base case earlier, $v \in \ker T^2 = \ker T$. Hence, by induction, $\ker T^k = \ker T$ for all positive integers k.
For the other equality, note that $v \in \operatorname{im} T^k$ implies that there exists $u \in V$ such that $v = T^k u$. Then $T(T^{k-1}u) = T^k u = v$, so $v \in \operatorname{im} T$. Hence, $\operatorname{im} T^k \subseteq \operatorname{im} T$. Conversely, note that

$$\dim \operatorname{im} T^k = \dim V - \dim \ker T^k = \dim V - \dim \ker T = \dim \operatorname{im} T$$

by the previous equality $\ker T^k = \ker T$. Since $\operatorname{im} T^k \subseteq \operatorname{im} T$ and they have the same dimension, we conclude that $\operatorname{im} T^k = \operatorname{im} T$. □

Solution 7A.28 If the minimal polynomial $p(x)$ of T is a polynomial multiple of $(x-\lambda)^2$, then $p(x) = (x-\lambda)^2 q(x)$ for some polynomial $q(x)$. Then $(T-\lambda I)^2 q(T) = 0$, but $(T-\lambda I)q(T) \neq 0$ since $p(x)$ is the minimal polynomial. Pick v such that $(T-\lambda I)q(T)v \neq 0$. Then $(T-\lambda I)^2 q(T)v = 0$, so $q(T)v \in \ker(T-\lambda I)^2$. By normality of T, we know that $(T-\lambda I)^* = T^* - \overline{\lambda} I$ which commutes with $T - \lambda I$. Hence, by Exercise 7A27, $\ker(T-\lambda I)^2 = \ker(T-\lambda I)$. Therefore, $q(T)v \in \ker(T-\lambda I)$, which implies that $(T-\lambda I)q(T)v = 0$. This contradicts our choice of v. Hence, the minimal polynomial of T cannot be a polynomial multiple of $(x-\lambda)^2$. □

Solution 7A.29 Let $V = \mathbb{F}^2$, and consider the standard orthonormal basis e_1, e_2. Define $T \in \operatorname{End}(V)$ by $T(e_1) = e_1 - e_2$ and $T(e_2) = e_1$. Then

$$\operatorname{Mat}(T) = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}, \quad \operatorname{Mat}(T^*) = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$$

so that $T^*e_1 = e_1 + e_2$ and $T^*e_2 = -e_1$. Moreover,

$$\|Te_1\|^2 = \|e_1 - e_2\|^2 = 2, \quad \|T^*e_1\|^2 = \|e_1 + e_2\|^2 = 2,$$

and

$$\|Te_2\|^2 = \|e_1\|^2 = 1, \quad \|T^*e_2\|^2 = \|-e_1\|^2 = 1.$$

Hence, $\|Te_k\| = \|T^*e_k\|$ for each $k = 1, 2$. However, T is not normal because

$$\operatorname{Mat}(TT^*) = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}, \quad \operatorname{Mat}(T^*T) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Therefore, $TT^* \neq T^*T$. □

Solution 7A.30 Since T is normal, we have $\ker T = \ker T^*$ by 7.20. Let $z = (z_1, z_2, z_3) \in \ker T = \ker T^*$. By $T(1,1,1) = 2(1,1,1)$ and the definition of the adjoint, we have

$$2\langle (1,1,1), z\rangle = \langle T(1,1,1), z\rangle = \langle (1,1,1), T^*z\rangle = \langle (1,1,1), 0\rangle = 0.$$

Then $\langle (1,1,1), z\rangle = 0$. Direct calculation shows that

$$\langle (1,1,1), z\rangle = \overline{z_1} + \overline{z_2} + \overline{z_3} = \overline{z_1 + z_2 + z_3}.$$

Since $\langle (1,1,1), z\rangle = 0$, we have $\overline{z_1 + z_2 + z_3} = 0$, which implies $z_1 + z_2 + z_3 = 0$. □

Solution 7A.31

(a) By a similar set of arguments as in Exercise 6B4, the list above is orthogonal with respect to the given inner product. Note that any $f, g \in V$ is a linear combination of $1, \cos kx, \sin kx$ for $k = 1, 2, \ldots, n$, all of which are periodic with period 2π. Then $f(\pi) = f(-\pi)$ and $g(\pi) = g(-\pi)$. Integration by parts gives

$$\langle Df, g\rangle = \int_{-\pi}^{\pi} f'g = fg\,\big|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} fg' = -\int_{-\pi}^{\pi} fg' = \langle f, -Dg\rangle.$$

Then $D^* = -D$, which shows that

$$DD^* = D(-D) = -D^2 \quad \text{and} \quad D^*D = (-D)D = -D^2.$$

Hence, $DD^* = D^*D$, which shows that D is normal. However, D is not self-adjoint, since $D\cos x = -\sin x \neq \sin x = D^* \cos x$.

(b) Note that $T = D^2$. By the previous part, $D^* = -D$, so $T^* = (D^2)^* = (D^*)^2 = (-D)^2 = D^2 = T$. Hence, T is self-adjoint.

□

Solution 7A.32 For all $v \in V$ and $w \in W$, we have

$$(T'(\phi_w))v = \phi_w(Tv) = \langle Tv, w\rangle = \langle v, T^*w\rangle = \phi_{T^*w}v.$$

Hence, $T'(\phi_w) = \phi_{T^*w}$ for all $w \in W$. □

§7B

Solution 7B.1 ⇒ This follows from 7.12.

⇐ Let T be a normal operator on a complex inner product space with only real eigenvalues. By the complex spectral theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of eigenvectors of T, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Because T is normal, each e_i is also an eigenvector of T^* with eigenvalue $\overline{\lambda_i}$ by 7.21(e). Since all the eigenvalues λ_i are real, we have $\overline{\lambda_i} = \lambda_i$ for all i. Hence,

$$Te_i = \lambda_i e_i = \overline{\lambda_i} e_i = T^* e_i$$

for all $i = 1, \ldots, n$. Since T and T^* agree on a basis of V, we have $T = T^*$ by the linear map lemma. □

Solution 7B.2 By the complex spectral theorem, T is diagonalizable. By 5.55, V is the direct sum of the eigenspaces of T. Since T has exactly one eigenvalue λ, $V = E(\lambda, T)$. Therefore, $Tv = \lambda v$ for all $v \in V$, which means $T = \lambda I$. □

Solution 7B.3 $\Rightarrow$ Suppose the set of eigenvalues of T is contained in $\{0, 1\}$. By the complex spectral theorem, T is diagonalizable. By 5.55, V is the direct sum of the eigenspaces of T. Thus, we can write $V = \ker(T - I) \oplus \ker T$. Let $U = \ker(T - I)$. By 7.22, $E(1, T)$ is orthogonal to $E(0, T)$, so $\ker T \subseteq U^\perp$. Since any $v \in V$ can be uniquely written as $v = u + w$ with $u \in U$ and $w \in U^\perp$, we have

$$Tv = T(u + w) = u + 0 = u = P_U v,$$

which shows that $T = P_U$.
$\Leftarrow$ If $T = P_U$, then by Exercise 7A20, we know $P_U^2 = P_U$. If $Tv = \lambda v$ for some $\lambda \in \mathbb{C}$ and $v \neq 0$, then $\lambda^2 v = \lambda v$. This yields $\lambda^2 = \lambda$, so $\lambda \in \{0, 1\}$. □

Solution 7B.4 $\Rightarrow$ Let $v \in V$ be an eigenvector of T with eigenvalue λ. From 7.21(e), we know that $Tv = \lambda v$ implies $T^* v = \overline{\lambda} v$. If T is skew, then $T = -T^*$, so $Tv = -T^* v = -\overline{\lambda} v$. Therefore, $\lambda = -\overline{\lambda}$, which means λ is purely imaginary.
$\Leftarrow$ Suppose all eigenvalues of T are purely imaginary. By the complex spectral theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Since T is normal, each e_i is also an eigenvector of T^* with eigenvalue $\overline{\lambda}_i$. Because each λ_i is purely imaginary, we have $\overline{\lambda_i} = -\lambda_i$ for each i. Therefore,

$$T^* e_i = \overline{\lambda}_i e_i = -\lambda_i e_i = -T e_i$$

for all $i = 1, \ldots, n$. Since T^* and $-T$ agree on the basis, $T^* = -T$ by the linear map lemma. □

Solution 7B.5 This is false. Consider $T \in \text{End}(\mathbb{C}^3)$ with matrix

$$\text{Mat}(T) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Since the diagonal entries are distinct, T is diagonalizable. However, $\|Te_1\| = 1$, while $\|T^* e_1\| = \sqrt{2}$, so T is not normal by 7.20. □

Solution 7B.6 By the complex spectral theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Define $Te_i = \lambda_i e_i$ for each $i = 1, \ldots, n$. From $T^9 = T^8$, we have $\lambda_i^9 = \lambda_i^8$, which implies $\lambda_i^8(\lambda_i - 1) = 0$. This gives $\lambda_i \in \{0, 1\}$. By Exercise 7B1, since all eigenvalues of T are real, T is self-adjoint. Moreover, since each eigenvalue λ_i satisfies $\lambda_i \in \{0, 1\}$, we have $\lambda_i^2 = \lambda_i$ for each i, so $T^2 e_i = \lambda_i^2 e_i = \lambda_i e_i = T e_i$ for each $i = 1, \ldots, n$. Since T^2 and T agree on the basis, $T^2 = T$ by the linear map lemma. □

Solution 7B.7 Consider $T \in \mathrm{End}(\mathbb{C}^2)$ defined by $T(z_1, z_2) = (z_2, 0)$. Then $T^2 = 0$, so $T^9 = T^8 = 0$, but $T^2 = 0 \neq T$. □

Solution 7B.8 $\Rightarrow$ This follows directly from 7.21(e).
$\Leftarrow$ If $V = \{0\}$, then T is trivially normal. Otherwise, assume $\dim V > 0$. By Schur's theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which T has an upper-triangular matrix A. We show by induction on k that $A_{k,j} = 0$ for all $j \neq k$, so that A is diagonal. Suppose $A_{i,j} = 0$ for all $j \neq i$ and all $i < k$. Since A is upper-triangular, $Te_k = \sum_{i=1}^{k} A_{i,k} e_i$. By the inductive hypothesis, $A_{i,k} = 0$ for all $i < k$. Hence, $Te_k = A_{k,k} e_k$, so e_k is an eigenvector of T. By hypothesis, e_k is an eigenvector of T^*, say $T^* e_k = \lambda e_k$. Since $\mathrm{Mat}(T^*) = A^*$, this gives

$$T^* e_k = \sum_{j=1}^{n} \overline{A_{k,j}} e_j = \lambda e_k,$$

and the uniqueness of the representation with respect to $e_1, \ldots, e_n$ forces $\overline{A_{k,j}} = 0$ for all $j \neq k$. Thus, A is diagonal, so A and A^* commute, and therefore $TT^* = T^*T$. □

Solution 7B.9 $\Rightarrow$ If $V = \{0\}$, then $T^* = 0 = p(T)$ trivially with $p = 0$. Otherwise, suppose $V \neq \{0\}$. If T is normal, then by the complex spectral theorem, there exists an orthonormal basis of V consisting of eigenvectors of T. Let $e_1, \ldots, e_n$ be such a basis, and let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Of the eigenvalues, let $\mu_1, \ldots, \mu_m$ be the distinct ones. By Exercise 4.7, there exists a polynomial $p \in \mathbb{C}[X]$ such that $p(\mu_i) = \overline{\mu_i}$ for all $i \in \{1, \ldots, m\}$. Since each eigenvalue λ_i is equal to some μ_j, we have $p(\lambda_i) = \overline{\lambda_i}$ for all i. Because T is normal and $Te_i = \lambda_i e_i$, we have $T^* e_i = \overline{\lambda_i} e_i$ by 7.21(e). Also, $p(T)e_i = p(\lambda_i)e_i = \overline{\lambda_i} e_i$. Hence, $T^* e_i = p(T)e_i$ for all i. Since T^* and $p(T)$ agree on a basis of V, we have $T^* = p(T)$ by the linear map lemma.
$\Leftarrow$ Suppose $T^* = p(T)$ for some polynomial $p \in \mathbb{C}[X]$. Since T commutes with $p(T)$, we have $TT^* = Tp(T) = p(T)T = T^*T$, so T is normal. □

Solution 7B.10

Axler Note. *An operator $S \in \mathrm{End}(V)$ is called a* square root *of $T \in \mathrm{End}(V)$ if $S^2 = T$. We will discuss more about square roots of operators in Sections 7C and 8C.*

Let $T \in \mathrm{End}(V)$ be normal. By the complex spectral theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. For each i, the fundamental theorem of algebra applied to the polynomial $z^2 - \lambda_i$ asserts the existence of $\mu_i \in \mathbb{C}$ such that $\mu_i^2 = \lambda_i$. Let $S \in \mathrm{End}(V)$ be the operator defined by $Se_i = \mu_i e_i$ for all i. Then

$$S^2 e_i = S(\mu_i e_i) = \mu_i S e_i = \mu_i(\mu_i e_i) = \mu_i^2 e_i = \lambda_i e_i = Te_i$$

for all i. Since S^2 and T agree on a basis of V, we have $S^2 = T$. □

Solution 7B.11

Axler Note. *An operator $S \in \mathrm{End}(V)$ is called a* cube root *of $T \in \mathrm{End}(V)$ if $S^3 = T$.*

Let $T \in \text{End}(V)$ be self-adjoint. By 7.29 (for $\mathbb{F} = \mathbb{R}$) or 7.31 (for $\mathbb{F} = \mathbb{C}$ by normality of T), there exists an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$. By 7.12, these eigenvalues $\lambda_1, \ldots, \lambda_n$ are real. For each i, the polynomial $z^3 - \lambda_i$ has a real root μ_i. Let $S \in \text{End}(V)$ be the operator defined by $Se_i = \mu_i e_i$ for all i. Then

$$S^3 e_i = S^2(\mu_i e_i) = S(\mu_i^2 e_i) = \mu_i^3 e_i = \lambda_i e_i = Te_i$$

for all i. Since S^3 and T agree on a basis of V, we have $S^3 = T$. □

Solution 7B.12

Manual Remark. *The original exercise statement said "Suppose V is a complex vector space...". I believe this to be an oversight, since T^* is defined using the inner product.*

By Exercise 7B9, we may write $T^* = p(T)$ for some polynomial p. Since S commutes with T, it also commutes with any polynomial in T, and hence with T^*. □

Solution 7B.13 Normality of T implies that T and T^* commute. By Exercise 6B20, there exists an orthonormal basis of V with respect to which both T and T^* have upper-triangular matrices. By 7.9, $\text{Mat}(T^*) = \text{Mat}(T)^*$. Since $\text{Mat}(T)^*$ is upper-triangular, $\text{Mat}(T)$ must be lower-triangular. This forces $\text{Mat}(T)$ to be diagonal. □

Solution 7B.14

Manual Remark. *This exercise adds another characterization to the real spectral theorem, where T is self-adjoint if and only if the eigenvectors corresponding to distinct eigenvalues are orthogonal and V decomposes as the direct sum of its eigenspaces. In particular, this is the orthogonal analogue of Theorem 5.55.*

Appendix. *An alternate proof is given at the end of this solution.*

$\Rightarrow$ Since T is self-adjoint, it is normal. By 7.22, its eigenvectors corresponding to distinct eigenvalues are orthogonal. By the real spectral theorem, V has an orthonormal basis consisting of eigenvectors of T, so T is diagonalizable. By 5.55, this gives the desired decomposition of V.

$\Leftarrow$ In each $E(\lambda_i, T)$, apply the Gram–Schmidt process to a basis of that eigenspace to obtain an orthonormal basis for that eigenspace. By hypothesis, these orthonormal bases for the eigenspaces corresponding to distinct eigenvalues are mutually orthogonal. The combined list of these orthonormal bases forms an orthonormal list of eigenvectors of T, hence is linearly independent by 6.25. Moreover, this list spans V because $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$ since each eigenspace is the span of its orthonormal basis obtained via Gram–Schmidt. Thus, the list is an orthonormal basis of V consisting of eigenvectors of T, so T is self-adjoint by the real spectral theorem.

Alternate proof. This solution rewrites the backwards direction of the proof, appealing to an orthogonal projection argument rather than the real spectral theorem.

Alternate proof. Civil Service Pigeon. Let P_i be the orthogonal projection onto $E(\lambda_i, T)$. By hypothesis, eigenvectors corresponding to distinct eigenvalues are orthogonal, so the eigenspaces are mutually orthogonal. Since

$$V = \bigoplus_{i=1}^{m} E(\lambda_i, T),$$

these eigenspaces orthogonally decompose V, yielding $\sum_{i=1}^{m} P_i = I$. For any $v \in V$,

$$Tv = T\left(\sum_{i=1}^{m} P_i v\right) = \sum_{i=1}^{m} \lambda_i P_i v \implies T = \sum_{i=1}^{m} \lambda_i P_i.$$

Since each P_i an orthogonal projection onto mutually orthogonal subspaces, $P_i^* = P_i$ and $P_i P_j = 0$ for $i \neq j$. □

Solution 7B.15

Manual Remark. *This has the same remark as Exercise 7B14 but in the context of the complex spectral theorem. It also uses normal operators instead of self-adjoint operators.*

Appendix. *An alternate proof is given at the end of this solution.*

The proof is similar to that of Exercise 7B14, but using the complex spectral theorem instead of the real spectral theorem. Moreover, the ⇒ direction applies 7.22 directly since T is normal by hypothesis.

Alternate proof. Civil Service Pigeon. Since $T^* = \sum_{i=1}^{m} \overline{\lambda_i} P_i$,

$$TT^* = \left(\sum_{i=1}^{m} \lambda_i P_i\right)\left(\sum_{j=1}^{m} \overline{\lambda_j} P_j\right) = \sum_{i=1}^{m} |\lambda_i|^2 P_i.$$

By symmetry, T^*T yields the same sum, so $TT^* = T^*T$ and T is normal. □

Solution 7B.16 ⇒ By assumption, there exists an orthonormal basis of V with respect to which every element of $\mathscr{E}$ has a diagonal matrix. For each $T \in \mathscr{E}$, the diagonal matrix $\mathrm{Mat}(T)$ commutes with its conjugate transpose $\mathrm{Mat}(T)^*$. By 7.9, we know $\mathrm{Mat}(T^*) = \mathrm{Mat}(T)^*$, so $\mathrm{Mat}(T)$ commutes with $\mathrm{Mat}(T^*)$, which implies that T is normal. By 3.43, for all $S, T \in \mathscr{E}$, we have

$$\mathrm{Mat}(ST) = \mathrm{Mat}(S)\,\mathrm{Mat}(T) = \mathrm{Mat}(T)\,\mathrm{Mat}(S) = \mathrm{Mat}(TS),$$

so $ST = TS$.

⇐ Taking $S = T$ in the hypothesis gives that each $T \in \mathscr{E}$ is normal. By Exercise 7B12 (Fuglede's theorem), for all $S, T \in \mathscr{E}$, $ST^* = T^*S$. Then

$$S^*T^* = (TS)^* = (ST)^* = T^*S^*,$$

so S^* and T^* commute. Letting $\mathscr{E}^* = \{T^* : T \in \mathscr{E}\}$, we see that $\mathscr{E} \cup \mathscr{E}^*$ is a set of commuting operators. By Exercise 6B20, there exists an orthonormal basis of V with respect to which every element of $\mathscr{E} \cup \mathscr{E}^*$ has an upper-triangular matrix. For each $T \in \mathscr{E}$, both $\mathrm{Mat}(T)$ and $\mathrm{Mat}(T^*)$ are upper-triangular, and $\mathrm{Mat}(T^*) = \mathrm{Mat}(T)^*$ by 7.9. Hence, $\mathrm{Mat}(T)$ is both upper-triangular and lower-triangular, which implies that $\mathrm{Mat}(T)$ is diagonal. Therefore, there exists an orthonormal basis of V with respect to which every element of $\mathscr{E}$ has a diagonal matrix. □

Solution 7B.17 $\Rightarrow$ Fix an orthonormal basis of V with respect to which every element of $\mathscr{C}$ has a diagonal matrix. For any $T \in \mathscr{C}$, the matrix of T with respect to this orthonormal basis is diagonal. Moreover, 7.9 shows that $\text{Mat}(T^*) = \text{Mat}(T)^* = \text{Mat}(T)$. Since the matrices are equal and over the same orthonormal basis, it follows that $T = T^*$, so T is self-adjoint. Additionally, for any $S, T \in \mathscr{C}$, the matrices of S and T with respect to this orthonormal basis are diagonal, and diagonal matrices commute. A similar argument from Exercise 7B16 shows that $ST = TS$.
$\Leftarrow$ We proceed by strong induction on $\dim V = n$. If $n = 0$, the empty list is an orthonormal basis of V, and the statement holds trivially. Now assume that the result holds for any real inner product space of dimension less than n and every collection of commuting self-adjoint operators on it. If every $T \in \mathscr{C}$ is some scalar multiple of the identity, then any orthonormal basis of V will diagonalize every element of $\mathscr{C}$, and the statement holds. In particular, this holds even with $\mathscr{C} = \varnothing$.

Otherwise, choose $T \in \mathscr{C}$ that is not a scalar multiple of the identity. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. By the real spectral theorem and Exercise 7B14, we may write

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where the eigenspaces are pairwise orthogonal. Note that if $m = 1$, then $V = E(\lambda_1, T)$ so that $T = \lambda_1 I$, contradicting our choice of T as not a scalar multiple of the identity. Therefore, $m \geq 2$. For any i, nontriviality of each eigenspace implies that $\dim E(\lambda_i, T) = n - \sum_{j \neq i} \dim E(\lambda_j, T) \leq n - 1$.

We now claim that each eigenspace $E(\lambda_i, T)$ is invariant under every $S \in \mathscr{C}$. To see this, let $v \in E(\lambda_i, T)$ and $S \in \mathscr{C}$. Then

$$T(Sv) = S(Tv) = S(\lambda_i v) = \lambda_i (Sv),$$

so $Sv \in E(\lambda_i, T)$. Hence, each eigenspace $E(\lambda_i, T)$ is invariant under every $S \in \mathscr{C}$.

Now consider each eigenspace under the inner product inherited from V, and let $\mathscr{C}_i = \{S|_{E(\lambda_i,T)} : S \in \mathscr{C}\}$. For any $u, w \in E(\lambda_i, T)$,

$$\langle S|_{E(\lambda_i,T)} u, w \rangle = \langle Su, w \rangle = \langle u, Sw \rangle = \langle u, S|_{E(\lambda_i,T)} w \rangle$$

so that each $S|_{E(\lambda_i,T)}$ is a self-adjoint operator on $E(\lambda_i, T)$. Furthermore, for any $S|_{E(\lambda_i,T)}, R|_{E(\lambda_i,T)} \in \mathscr{C}_i$, we have

$$S|_{E(\lambda_i,T)} R|_{E(\lambda_i,T)} = (SR)|_{E(\lambda_i,T)} = (RS)|_{E(\lambda_i,T)} = R|_{E(\lambda_i,T)} S|_{E(\lambda_i,T)},$$

so $\mathscr{C}_i$ is a commuting family of self-adjoint operators on $E(\lambda_i, T)$. Then the induction hypothesis applies to $E(\lambda_i, T)$ and $\mathscr{C}_i$, so there exists an orthonormal basis of $E(\lambda_i, T)$ with respect to which every element of $\mathscr{C}_i$ has a diagonal matrix. By Exercise 7B14, concatenating these bases gives an orthonormal basis of V since the eigenspaces are pairwise orthogonal, and V is the direct sum of these eigenspaces. Now let $S \in \mathscr{C}$. For each basis vector e in this orthonormal basis, there exists an i such that e is one of the basis vectors of $E(\lambda_i, T)$. Since $\text{Mat}(S|_{E(\lambda_i,T)})$ is diagonal with respect to the chosen orthonormal basis of $E(\lambda_i, T)$, it follows that $Se = \mu_e e$ for some $\mu_e \in \mathbb{R}$. Therefore, the matrix of S with respect to this orthonormal basis is diagonal. $\square$

Solution 7B.18 Consider the real inner product space $V = \mathbb{R}^2$ and the operator $T \in \mathrm{End}(V)$ whose matrix with respect to the standard basis is

$$\mathrm{Mat}(T) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

so that $T^2 = -I$. Now choose $b = 0$ and $c = 1$. Then $b^2 = 0 < 4c = 4$, and

$$T^2 + bT + cI = T^2 + I = -I + I = 0,$$

which is not invertible. □

Solution 7B.19

(a) For any $u \in U$ and $w \in U^\perp$, we have

$$\langle Tw, u\rangle = \langle w, Tu\rangle = 0$$

since $Tu \in U$. Hence, $Tw \in U^\perp$.

(b) For any $u_1, u_2 \in U$, we have

$$\langle T|_U u_1, u_2\rangle = \langle Tu_1, u_2\rangle = \langle u_1, Tu_2\rangle = \langle u_1, T|_U u_2\rangle.$$

Hence, $T|_U$ is self-adjoint.

(c) Applying the same argument as in part (b) to $U^\perp$ gives the conclusion.

□

Solution 7B.20

Manual Remark. *The ordering of the parts in this exercise is strange. The solution below will prove (b) first as (a) falls out naturally as a consequence.*

(b) If $U = \{0\}$, the result is immediate since the zero subspace is invariant under any operator. Otherwise, let $e_1, \ldots, e_n$ be an orthonormal basis of U. By definition of P_U, we have for all $v \in V$ that $P_U v = \sum_{i=1}^n \langle v, e_i\rangle e_i$. Moreover, the Pythagorean theorem gives $\|v\|^2 = \|P_U v\|^2 + \|v - P_U v\|^2$. Hence, $\|v\|^2 = \sum_{i=1}^n |\langle v, e_i\rangle|^2$ if and only if $v \in U$.

Now, for each $j = 1, \ldots, n$, the invariance of U under T gives $Te_j \in U$. Hence,

$$\|Te_j\|^2 = \sum_{i=1}^n |\langle Te_j, e_i\rangle|^2.$$

By Bessel's inequality, we have for all $j = 1, \ldots, n$ that

$$\|T^* e_j\|^2 \geq \sum_{i=1}^n |\langle T^* e_j, e_i\rangle|^2 = \sum_{i=1}^n |\langle e_j, Te_i\rangle|^2.$$

Since T is normal, we have $\|T^* e_j\| = \|Te_j\|$ for each $j = 1, \ldots, n$. Summing over $j = 1, \ldots, n$ gives

$$\sum_{j=1}^n \|Te_j\|^2 = \sum_{j=1}^n \|T^* e_j\|^2 \geq \sum_{j=1}^n \sum_{i=1}^n |\langle T^* e_j, e_i\rangle|^2 = \sum_{i=1}^n \sum_{j=1}^n |\langle e_j, Te_i\rangle|^2 = \sum_{i=1}^n \|Te_i\|^2,$$

where $Te_i \in U$ by the invariance of U under T. Since the first and last sums in the inequality are equal, equality must hold throughout. In particular,

$$\sum_{j=1}^{n}\left(\|T^*e_j\|^2 - \sum_{i=1}^{n}|\langle T^*e_j, e_i\rangle|^2\right) = 0.$$

By Bessel's inequality, each term in the sum is nonnegative. Hence, each term must be zero, and

$$\|T^*e_j\|^2 - \sum_{i=1}^{n}|\langle T^*e_j, e_i\rangle|^2 = 0 \quad \text{for each } j = 1, \ldots, n.$$

The above equality shows that $T^*e_j \in U$ for each $j = 1, \ldots, n$. The linearity of T^* then implies that $T^*u \in U$ for all $u \in U$, so U is invariant under T^*.

(a) Let $w \in U^\perp$ and $u \in U$. Using (b), $T^*u \in U$, so $\langle w, T^*u\rangle = \langle Tw, u\rangle = 0$. Hence, $Tw \in U^\perp$, and $U^\perp$ is invariant under T.

(c) By (b), we know $(T^*)|_U \in \text{End}(U)$. Then for all $x, y \in U$, we have

$$\langle (T|_U)y, x\rangle = \langle Ty, x\rangle = \langle y, T^*x\rangle = \langle y, (T^*)|_U x\rangle.$$

Since the adjoint is unique, we conclude that $(T|_U)^* = (T^*)|_U$.

(d) For any $u' \in U$, (b) gives $T^*u' \in U$. Using (c), we have

$$(T|_U)(T|_U)^*u' = T(T^*u') = (TT^*)u' = (T^*T)u' = T^*(Tu') = (T|_U)^*(T|_U)u',$$

showing normality of $T|_U$.

Now, by (a), we have that $U^\perp$ is invariant under T. Then we may use the arguments in (b) and (c) with $U^\perp$ in place of U to conclude that $U^\perp$ is invariant under T^*. Additionally, the argument of the first paragraph with $U^\perp$ in place of U shows that $T|_{U^\perp}$ is normal.

□

Solution 7B.21 By either the real or complex spectral theorem, we may take an orthonormal basis of eigenvectors $e_1, \ldots, e_n$ of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. For each $i \in \{1, \ldots, n\}$, we have $\lambda_i \in \{2, 3\}$. Then

$$(T^2 - 5T + 6I)e_i = (\lambda_i^2 - 5\lambda_i + 6)e_i = (\lambda_i - 2)(\lambda_i - 3)e_i = 0.$$

Since $T^2 - 5T + 6I$ evaluates to 0 on each basis vector e_i, the conclusion follows. □

Solution 7B.22 Consider $T \in \text{End}(\mathbb{C}^3)$ with matrix representation

$$\text{Mat}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Since T is upper-triangular with diagonal entries $2, 2, 3$, these are the eigenvalues of T. However,

$$(T - 2I)(T - 3I)e_2 = (T - 2I)(e_1 - e_2) = -e_1 \neq 0.$$ □

Solution 7B.23 If $\mathbb{F} = \mathbb{R}$, the real spectral theorem guarantees an orthonormal basis of eigenvectors. If $\mathbb{F} = \mathbb{C}$, then T is normal, since T is self-adjoint, and the complex spectral theorem guarantees an orthonormal basis of eigenvectors. In either case, we obtain an orthonormal basis of eigenvectors $e_1, \ldots, e_n$ of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Hence, we may write $v = \sum_{i=1}^n \alpha_i e_i$ for scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Since the basis is orthonormal and $\|v\| = 1$, we have $\sum_{i=1}^n |\alpha_i|^2 = 1$. Then

$$Tv - \lambda v = \sum_{i=1}^n \alpha_i \lambda_i e_i - \sum_{i=1}^n \alpha_i \lambda e_i = \sum_{i=1}^n \alpha_i (\lambda_i - \lambda) e_i.$$

By orthonormality of the basis, we have

$$\|Tv - \lambda v\|^2 = \sum_{i=1}^n |\alpha_i|^2 |\lambda - \lambda_i|^2.$$

Now, choose k such that

$$|\lambda - \lambda_k| = \min_{1 \le i \le n} |\lambda - \lambda_i|.$$

Let $\lambda' = \lambda_k$ so that λ' is an eigenvalue of T. Since $|\lambda - \lambda'| = |\lambda - \lambda_k| \le |\lambda - \lambda_i|$ for all i, we have

$$|\lambda - \lambda'|^2 = |\lambda - \lambda_k|^2 \sum_{i=1}^n |\alpha_i|^2 \le \sum_{i=1}^n |\alpha_i|^2 |\lambda - \lambda_i|^2 = \|Tv - \lambda v\|^2 < \varepsilon^2.$$

Taking square roots, we conclude that $|\lambda - \lambda'| < \varepsilon$. □

Solution 7B.24 Since U is only assumed to be a finite-dimensional vector space, we require an inner product to discuss orthonormality in the $\Leftarrow$ direction of both parts. Hence, we prove the following lemma: there exists an inner product on U that makes any basis $e_1, \ldots, e_n$ orthonormal.
Let $f_1, \ldots, f_n$ denote the standard basis of $\mathbb{F}^n$, and equip $\mathbb{F}^n$ with the Euclidean inner product. Define $\phi \in \mathrm{Hom}(U, \mathbb{F}^n)$ to be the unique linear map

$$\phi(e_i) = f_i \quad \text{for } i = 1, \ldots, n.$$

Then for any $u \in U$ and scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $u = \sum_{i=1}^n \alpha_i e_i$, we have $\phi(u) = \sum_{i=1}^n \alpha_i f_i$. Additionally, if $\phi(u) = 0$, then $\alpha_i = 0$ for each i so that $u = 0$. Hence, ϕ is injective. As in Exercise 6A2, applied verbatim to a linear map rather than a linear operator, we may define an inner product on U as

$$\langle u, v \rangle = \langle \phi(u), \phi(v) \rangle \quad \text{for all } u, v \in U.$$

Moreover, $\langle e_i, e_j \rangle = \langle f_i, f_j \rangle = \delta_{ij}$ for all i, j, so the basis $e_1, \ldots, e_n$ is orthonormal with respect to this inner product. This completes the proof of the lemma.

(a) $\Rightarrow$ Suppose T is diagonalizable. Then the matrix of T with respect to some basis $u_1, \ldots, u_n$ consisting of eigenvectors of T is a diagonal matrix whose diagonal entries are eigenvalues of T. Since a diagonal matrix is equal to its transpose, the matrix of T with respect to the basis $u_1, \ldots, u_n$ equals its transpose.

$\Leftarrow$ Suppose there is a basis $e'_1, \dots, e'_n$ of U such that the matrix of T with respect to $e'_1, \dots, e'_n$ equals its transpose. By the lemma, there exists an inner product on U that makes the basis $e'_1, \dots, e'_n$ orthonormal. Let $\mathrm{Mat}(T)$ denote the matrix of T with respect to the basis $e'_1, \dots, e'_n$; similarly for $\mathrm{Mat}(T^*)$. By 7.9, $\mathrm{Mat}(T^*) = \mathrm{Mat}(T)^*$, and since all entries of $\mathrm{Mat}(T)$ are real, $\mathrm{Mat}(T)^* = \mathrm{Mat}(T)^t$. Moreover, the hypothesis gives $\mathrm{Mat}(T)^t = \mathrm{Mat}(T)$. Therefore, $\mathrm{Mat}(T^*) = \mathrm{Mat}(T)$ so that T is self-adjoint. By the real spectral theorem, T has a diagonal matrix with respect to some orthonormal basis, hence T is diagonalizable.

(b) $\Rightarrow$ Suppose T is diagonalizable. Then there exists a basis $v_1, \dots, v_n$ of U consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$. By definition, $Tv_i = \lambda_i v_i$ for each $i = 1, \dots, n$. Let $A = \mathrm{Mat}(T, (v_1, \dots, v_n))$ so that A is a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. It follows that A^* is also a diagonal matrix with diagonal entries $\overline{\lambda_1}, \dots, \overline{\lambda_n}$. Since diagonal matrices commute, we have $AA^* = A^*A$.

$\Leftarrow$ Suppose there is a basis $e''_1, \dots, e''_n$ of U such that the matrix of T with respect to $e''_1, \dots, e''_n$ commutes with its conjugate transpose. By the lemma, there exists an inner product on U that makes the basis $e''_1, \dots, e''_n$ orthonormal. For any $S \in \mathrm{End}(U)$, let $\mathrm{Mat}(S)$ denote the matrix of S with respect to the basis $e''_1, \dots, e''_n$. By 7.9, $\mathrm{Mat}(T^*) = \mathrm{Mat}(T)^*$, and the hypothesis gives $\mathrm{Mat}(T)\,\mathrm{Mat}(T)^* = \mathrm{Mat}(T)^*\,\mathrm{Mat}(T)$. By 3.43, we have

$$\mathrm{Mat}(TT^*) = \mathrm{Mat}(T)\,\mathrm{Mat}(T^*) = \mathrm{Mat}(T)\,\mathrm{Mat}(T)^* = \mathrm{Mat}(T)^*\,\mathrm{Mat}(T) = \mathrm{Mat}(T^*)\,\mathrm{Mat}(T) = \mathrm{Mat}(T^*T),$$

so $TT^* = T^*T$, and hence T is normal. By the complex spectral theorem, T has a diagonal matrix with respect to some orthonormal basis of U, hence T is diagonalizable. □

Solution 7B.25 Let $S = \{j : \lambda_j \neq 0\}$. For any $v \in V$, we can write

$$v = \sum_{i=1}^{n} \langle v, e_i \rangle e_i.$$

Then

$$Tv = \sum_{i=1}^{n} \langle v, e_i \rangle Te_i = \sum_{i=1}^{n} \langle v, e_i \rangle \lambda_i e_i.$$

It follows that $\ker T = \mathrm{span}(e_j : j \notin S)$, and $\mathrm{im}\, T = \mathrm{span}(e_i : i \in S)$. If $k \notin S$, then e_k is orthogonal to $\mathrm{im}\, T$, so $P_{\mathrm{im}\, T} e_k = 0$, and hence $T^\dagger e_k = 0$. If $k \in S$, then $e_k \in \mathrm{im}\, T$ so that $P_{\mathrm{im}\, T} e_k = e_k$. Since $\langle e_k, e_j \rangle = 0$ for all $j \notin S$, and these vectors span $\ker T$, we have $\frac{1}{\lambda_k} e_k \in (\ker T)^\perp$. Since $T(\frac{1}{\lambda_k} e_k) = e_k$, we have that $(T|_{(\ker T)^\perp})^{-1} e_k = \frac{1}{\lambda_k} e_k$, and hence $T^\dagger e_k = \frac{1}{\lambda_k} e_k$. This is exactly the formula we wanted to show. □

§7C

Solution 7C.1 For every $v \in V$, we have $\langle Tv, v\rangle \geq 0$ and $\langle -Tv, v\rangle = -\langle Tv, v\rangle \geq 0$. Both inequalities force $\langle Tv, v\rangle = 0$ for every $v \in V$. By 7.43, $Tv = 0$ for every $v \in V$, so $T = 0$. □

Solution 7C.2 Note that the standard basis is orthonormal, and $\mathrm{Mat}(T)$ as given is equal to its conjugate transpose, so T is self-adjoint. For any $x = (x_1, x_2, x_3, x_4) \in \mathbb{F}^4$, we have

$$Tx = (2x_1 - x_2, -x_1 + 2x_2 - x_3, -x_2 + 2x_3 - x_4, -x_3 + 2x_4).$$

Since $x_i\overline{x_j} + \overline{x_i}x_j = 2\,\mathrm{Re}(x_i\overline{x_j})$, we have

$$\begin{aligned}\langle Tx, x\rangle &= (2x_1 - x_2)\overline{x_1} + (-x_1 + 2x_2 - x_3)\overline{x_2} + (-x_2 + 2x_3 - x_4)\overline{x_3} + (-x_3 + 2x_4)\overline{x_4}\\ &= 2|x_1|^2 - \overline{x_1}x_2 - x_1\overline{x_2} + 2|x_2|^2 - \overline{x_2}x_3 - x_2\overline{x_3} + 2|x_3|^2 - \overline{x_3}x_4 - x_3\overline{x_4} + 2|x_4|^2\\ &= 2|x_1|^2 - 2\,\mathrm{Re}(\overline{x_1}x_2) + 2|x_2|^2 - 2\,\mathrm{Re}(\overline{x_2}x_3) + 2|x_3|^2 - 2\,\mathrm{Re}(\overline{x_3}x_4) + 2|x_4|^2\\ &= |x_1|^2 + |x_1 - x_2|^2 + |x_2 - x_3|^2 + |x_3 - x_4|^2 + |x_4|^2 \geq 0.\end{aligned}$$

Hence, T is a positive operator. Moreover, if $Tx = 0$ for some $x \in \mathbb{F}^4$, then

$$0 = \langle Tx, x\rangle = |x_1|^2 + |x_1 - x_2|^2 + |x_2 - x_3|^2 + |x_3 - x_4|^2 + |x_4|^2.$$

This forces $x_1 = x_4 = 0$ so that $x_2 = x_3 = 0$ as well. Hence, $x = 0$, and therefore T is invertible by 3.65. □

Solution 7C.3 Since the standard basis is orthonormal, $\mathrm{Mat}(T)$ is equal to its conjugate transpose, so T is self-adjoint by 7.9. Let $x = (1, \ldots, 1) \in \mathbb{F}^n$. For any $y \in \mathbb{F}^n$, observe that $Ty = \langle y, x\rangle x$. Therefore,

$$\langle Ty, y\rangle = \langle\langle y, x\rangle x, y\rangle = \langle y, x\rangle\langle x, y\rangle = |\langle y, x\rangle|^2 \geq 0$$

for all $y \in \mathbb{F}^n$. Hence, T is a positive operator. □

Solution 7C.4 Let A be the n-by-n matrix whose entries are all 1 except $A_{1,2} = A_{2,1} = 2$. Then all entries of A are positive numbers, and $A = A^*$ by construction. Let $T \in \mathrm{End}(\mathbb{F}^n)$ be the operator whose matrix with respect to the standard basis is A. However, consider $x = (1, -1, 0, \ldots, 0) \in \mathbb{F}^n$. Then $(Tx)_1 = 1 + 2(-1) = -1$, $(Tx)_2 = 2(1) + (-1) = 1$, and $(Tx)_i = 1 - 1 = 0$ for $i > 2$. Hence,

$$\langle Tx, x\rangle = (-1)(1) + (1)(-1) + 0 = -2$$

so that T is not a positive operator. □

Solution 7C.5 Note that the (j, k)th entry of $\mathrm{Mat}(T, (e_1, \ldots, e_n))$ is $\langle Te_k, e_j\rangle$, hence the diagonal entries are $\langle Te_j, e_j\rangle$ for $j = 1, \ldots, n$.
$\Rightarrow$ Since T is self-adjoint by hypothesis, T is positive if and only if $\langle Tv, v\rangle \geq 0$ for all $v \in V$. In particular, this must hold for each $v = e_j$ in any orthonormal basis $e_1, \ldots, e_n$ of V, which shows that all diagonal entries of $\mathrm{Mat}(T, (e_1, \ldots, e_n))$ are nonnegative.

$\Leftarrow$ If $v = 0$, then $\langle Tv, v\rangle = 0 \geq 0$. Suppose $v \neq 0$, and let $e_1 = v/\|v\|$. Then $\|e_1\| = 1$, so e_1 is an orthonormal list by itself. Extend e_1 to an orthonormal basis

$e_1, \ldots, e_n$ of V. Then the first diagonal entry of $\mathrm{Mat}(T, (e_1, \ldots, e_n))$ is $\langle Te_1, e_1\rangle = \langle T(v/\|v\|), v/\|v\|\rangle = \langle Tv, v\rangle/\|v\|^2 \geq 0$. Hence, $\langle Tv, v\rangle \geq 0$ for all $v \in V$, which shows that T is a positive operator. □

Solution 7C.6 Let $S, T \in \mathrm{End}(V)$ be positive operators. In particular, S and T are self-adjoint. By 7.5(a), $S + T$ is also self-adjoint. Positivity follows from the fact that for any $v \in V$,

$$\langle (S + T)v, v\rangle = \langle Sv, v\rangle + \langle Tv, v\rangle \geq 0 + 0 = 0,$$

since S and T are positive operators. Hence, $S + T$ is a positive operator. □

Solution 7C.7 If $(S+T)v = 0$ for some $v \in V$, then $\langle Sv, v\rangle + \langle Tv, v\rangle = \langle (S+T)v, v\rangle = \langle 0, v\rangle = 0$. Since S and T are positive, $\langle Sv, v\rangle \geq 0$ and $\langle Tv, v\rangle \geq 0$. In particular, $\langle Sv, v\rangle = 0$. By 7.43, $Sv = 0$ so that $v = 0$ by the invertibility of S. Therefore, $(S + T)v = 0$ only if $v = 0$, which shows that $S + T$ is invertible by 3.65. □

Solution 7C.8 $\Rightarrow$ By 7.38(c), there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T with respect to which the matrix of T is diagonal with nonnegative diagonal entries $\lambda_1, \ldots, \lambda_n$. Hence, $Te_j = \lambda_j e_j$, where $\lambda_j \geq 0$ for each j. By Exercise 7B25,

$$T^\dagger e_j = \begin{cases} \frac{1}{\lambda_j} e_j & \text{if } \lambda_j \neq 0, \\ 0 & \text{if } \lambda_j = 0. \end{cases}$$

Hence, $\mathrm{Mat}(T^\dagger, (e_1, \ldots, e_n))$ is a diagonal matrix with nonnegative entries, which shows that $T^\dagger$ is a positive operator by 7.38(c).
$\Leftarrow$ By Exercise 6C23, $(T^\dagger)^\dagger = T$. Applying the forward direction to $T^\dagger$ in place of T shows that T is positive. □

Solution 7C.9 Note that $S^* \in \mathrm{Hom}(V, W)$, so $S^*TS \in \mathrm{End}(W)$. By 7.5(c) and 7.5(d), we have $(S^*TS)^* = S^*T^*(S^*)^* = S^*TS$, where the last equality holds since T is self-adjoint. Moreover, for any $w \in W$,

$$\langle S^*TSw, w\rangle = \langle TSw, Sw\rangle \geq 0,$$

since T is positive. Hence, S^*TS is a positive operator on W. □

Solution 7C.10

Manual Remark. *Alternatively, observe that* $T(v - w) = -(v - w)$, *so* $v - w \in E(-1, T) = \{0\}$ *by 7.38(b) and the fact that* T *is positive. Then* $v - w = 0$, *so* $v = w$.

Consider $v - w$. Then

$$\langle T(v-w), v-w\rangle = \langle Tv-Tw, v-w\rangle = \langle w-v, v-w\rangle = -\langle v-w, v-w\rangle = -\|v-w\|^2.$$

Since T is positive, $\langle T(v-w), v-w\rangle \geq 0$, so $-\|v-w\|^2 \geq 0$, which implies $\|v-w\|^2 = 0$ and hence $v - w = 0$, i.e., $v = w$. □

Solution 7C.11 Note that $T|_U \in \text{End}(U)$ since U is invariant under T, and U carries the inner product inherited from V. Since T is self-adjoint on V, Exercise 7B19(b) gives that $T|_U$ is self-adjoint on U. Moreover, for any $u \in U$,

$$\langle T|_U u, u\rangle_U = \langle Tu, u\rangle_V \geq 0,$$

since T is positive on V. Hence, $T|_U$ is a positive operator on U. □

Solution 7C.12 By 7.38(c), there is an orthonormal basis $e_1, \ldots, e_n$ of V with respect to which $\text{Mat}(T, (e_1, \ldots, e_n))$ is diagonal with nonnegative entries $\lambda_1, \ldots, \lambda_n$. By 3.43 and induction on k, $\text{Mat}(T^k, (e_1, \ldots, e_n)) = (\text{Mat}(T, (e_1, \ldots, e_n)))^k$ is diagonal with entries $\lambda_1^k, \ldots, \lambda_n^k$, which are all nonnegative. By 7.38(c) again, this shows that T^k is positive. □

Solution 7C.13

(a) $\Rightarrow$ If $T - \alpha I$ is positive, then its eigenvalues are nonnegative by 7.38(b). Since $Tv = \lambda v$ if and only if $(T - \alpha I)v = (\lambda - \alpha)v$, the eigenvalues of $T - \alpha I$ are precisely $\lambda - \alpha$, where λ ranges over the eigenvalues of T. Hence, $\alpha \leq \lambda$ for every eigenvalue λ of T.

$\Leftarrow$ By 7.5(a), (b), (e), self-adjointness of T, and $\alpha \in \mathbb{R}$,

$$(T - \alpha I)^* = T^* - \overline{\alpha} I^* = T - \alpha I,$$

so $T - \alpha I$ is self-adjoint. Since $\lambda - \alpha \geq 0$ for every eigenvalue λ of T, the eigenvalues of $T - \alpha I$ are nonnegative. Hence, $T - \alpha I$ is positive by 7.38(b).

(b) Note that $-T$ is self-adjoint by 7.5(b), and its eigenvalues are exactly $-\lambda$, where λ ranges over the eigenvalues of T. Applying (a) to $-T$ and $-\alpha$, the operator $-T - (-\alpha)I = \alpha I - T$ is positive if and only if $-\alpha \leq -\lambda$, i.e., $\alpha \geq \lambda$, for every eigenvalue λ of T.

□

Solution 7C.14 Let $v = \sum_{k=1}^m v_k$. By linearity of T, additivity of the inner product, and positivity of T, we have

$$\sum_{j=1}^m \sum_{k=1}^m \langle Tv_k, v_j\rangle = \sum_{j=1}^m \langle Tv, v_j\rangle = \langle Tv, v\rangle \geq 0.$$ □

Solution 7C.15 If $\mathbb{F} = \mathbb{R}$, the real spectral theorem applies directly, and if $\mathbb{F} = \mathbb{C}$, self-adjointness implies normality, so the complex spectral theorem applies. In either case, there exists an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. Since T is self-adjoint, $T^*T = T^2$. By the linear map lemma, there exists an operator C on V such that $Ce_k = |\lambda_k| e_k$ for each $k = 1, \ldots, n$. Since $\text{Mat}(C, (e_1, \ldots, e_n))$ is diagonal with nonnegative entries, C is a positive operator by 7.38(c). Moreover, $C^2 e_k = |\lambda_k|^2 e_k = \lambda_k^2 e_k = T^2 e_k$ for each k, so $C^2 = T^2$ by the linear map lemma. By 7.39, $C = \sqrt{T^2} = \sqrt{T^*T}$.

Consider now the operators

$$A = \frac{1}{2}(C + T) \quad \text{and} \quad B = \frac{1}{2}(C - T).$$

By construction, $A - B = T$ and $A + B = C = \sqrt{T^*T}$. Then for each $k = 1, \dots, n$,

$$Ae_k = \frac{1}{2}(C + T)e_k = \frac{1}{2}(|\lambda_k| + \lambda_k)e_k \quad \text{and} \quad Be_k = \frac{1}{2}(C - T)e_k = \frac{1}{2}(|\lambda_k| - \lambda_k)e_k.$$

Then the diagonal entries of $\operatorname{Mat}(A, (e_1, \dots, e_n))$ and $\operatorname{Mat}(B, (e_1, \dots, e_n))$ are nonnegative, so A and B are positive operators by 7.38(c). Finally,

$$ABe_k = \frac{1}{4}((C + T)(C - T))e_k = \frac{1}{4}((|\lambda_k| + \lambda_k)(|\lambda_k| - \lambda_k))e_k = 0 \quad \text{and} \quad BAe_k = 0$$

so that $AB = BA = 0$ by the linear map lemma. □

Solution 7C.16 Since T is positive, $\sqrt{T}$ is also a positive operator. In particular, it is self-adjoint, hence normal. Let $S = \sqrt{T}$, so $S^2 = T$. By Exercise 7A27, $\ker S = \ker S^2 = \ker T$ and $\operatorname{im} S = \operatorname{im} S^2 = \operatorname{im} T$. □

Solution 7C.17 If $V = \{0\}$, take $p = 0$. Otherwise, $V \neq \{0\}$. By 7.38(c), there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $Te_k = \lambda_k e_k$ where $\lambda_k \geq 0$ for each $k = 1, \dots, n$. Since $n \geq 1$, T has at least one eigenvalue. Let $\mu_1, \dots, \mu_m$ be the distinct eigenvalues of T, so $m \geq 1$. Because $e_k \neq 0$, λ_k is an eigenvalue of T so that $\lambda_k \in \{\mu_1, \dots, \mu_m\}$ for each $k = 1, \dots, n$. By 7.38(b), each $\mu_j \geq 0$ so that $\sqrt{\mu_j} \in \mathbb{R}$ for each $j = 1, \dots, m$. By Exercise 4.7 applied over $\mathbb{R}$, there exists $p \in \mathbb{R}[X]_{\leq m-1}$ such that $p(\mu_j) = \sqrt{\mu_j}$ for each $j = 1, \dots, m$. Since $T^i e_k = \lambda_k^i e_k$ for each nonnegative integer i, the definition of $p(T)$ gives $p(T)e_k = p(\lambda_k)e_k$ for each $k = 1, \dots, n$. Since $\lambda_k \in \{\mu_1, \dots, \mu_m\}$ and $p(\mu_j) = \sqrt{\mu_j}$ for each $j = 1, \dots, m$, we have $p(\lambda_k) = \sqrt{\lambda_k}$ for each $k = 1, \dots, n$. Then $\operatorname{Mat}(p(T), (e_1, \dots, e_n))$ is diagonal with diagonal entries $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$. By 7.38(c), $p(T)$ is positive. Then

$$p(T)^2 e_k = p(\lambda_k)^2 e_k = (\sqrt{\lambda_k})^2 e_k = \lambda_k e_k = Te_k$$

for each $k = 1, \dots, n$. By the linear map lemma, $p(T)^2 = T$, and by 7.39, $p(T) = \sqrt{T}$. □

Solution 7C.18 $\Rightarrow$ Note that S, T, and ST are all self-adjoint. By Exercise 7A9, ST is self-adjoint if and only if $ST = TS$.

$\Leftarrow$ By Exercise 7C17, there exists a polynomial p such that $\sqrt{S} = p(S)$. Since T commutes with S, it commutes with powers of S so that T commutes with $p(S) = \sqrt{S}$. Then $ST = \sqrt{S}\sqrt{S}T = \sqrt{S}^*\sqrt{S}T = \sqrt{S}^*T\sqrt{S}$, where the second equality uses the fact that $\sqrt{S}$ is self-adjoint. By Exercise 7C9, ST is positive. □

Solution 7C.19 Let $x, y \in \mathbb{R}$, and define the family of operators $T_{x,y}$ on $\mathbb{F}^2$ whose matrices with respect to the standard basis are

$$\operatorname{Mat}(T_{x,y}) = \begin{pmatrix} x & y \\ y & -x \end{pmatrix}.$$

Then $\text{Mat}(T_{x,y}) = \text{Mat}(T_{x,y})^*$. Since the standard basis is orthonormal, $T_{x,y}$ is self-adjoint by 7.9. Moreover,

$$\text{Mat}(T_{x,y})^2 = \begin{pmatrix} x & y \\ y & -x \end{pmatrix}^2 = \begin{pmatrix} x^2+y^2 & 0 \\ 0 & x^2+y^2 \end{pmatrix} = (x^2+y^2)I.$$

Taking $x = \cos\theta$ and $y = \sin\theta$ for $\theta \in \mathbb{R}$, we get $x^2+y^2 = 1$. Since $(\cos\theta, \sin\theta)$ takes on infinitely many distinct values and $T_{\cos\theta,\sin\theta}(1,0) = (\cos\theta, \sin\theta)$, the operators $T_{\cos\theta,\sin\theta}$ are all distinct. Therefore, the identity operator on $\mathbb{F}^2$ has infinitely many self-adjoint square roots. □

Solution 7C.20

Axler Note. *The numbers $\{\langle Te_k, e_j\rangle\}_{j,k=1,\ldots,n}$ are the entries in the matrix of T with respect to the orthonormal basis $e_1, \ldots, e_n$.*

$\Rightarrow$ By 7.38(f), let $T = R^*R$ for some $R \in \text{End}(V)$. Define $v_k = Re_k$ for $k = 1, \ldots, n$. Then for all $j, k = 1, \ldots, n$, we have

$$\langle v_k, v_j\rangle = \langle Re_k, Re_j\rangle = \langle R^*Re_k, e_j\rangle = \langle Te_k, e_j\rangle.$$

$\Leftarrow$ By the linear map lemma, we may define the unique operator $R \in \text{End}(V)$ such that $Re_k = v_k$ for $k = 1, \ldots, n$. Then

$$\langle Te_k, e_j\rangle = \langle v_k, v_j\rangle = \langle Re_k, Re_j\rangle = \langle R^*Re_k, e_j\rangle.$$

Let $S = T - R^*R$. Then

$$\langle Se_k, e_j\rangle = 0$$

for all $j, k = 1, \ldots, n$. By 6.30(a), we have

$$Se_k = \sum_{j=1}^{n} \langle Se_k, e_j\rangle e_j = 0.$$

Since this holds for all $k = 1, \ldots, n$, we have $S = 0$ by the linear map lemma so that $T = R^*R$. Therefore, T is positive by 7.38(f). □

Solution 7C.21

Axler Note. *Example: The 4-by-4 Hilbert matrix is*

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}.$$

Let $e_1, \ldots, e_n$ be the orthonormal basis of V with respect to which the matrix of T is the n-by-n Hilbert matrix. Let $H_n = \text{Mat}(T, (e_1, \ldots, e_n))$. Since H_n has real entries, and $\frac{1}{j+k-1} = \frac{1}{k+j-1}$, it follows that $H_n = H_n^*$. By 7.9, it follows that T is self-adjoint.

Now, observe that

$$\int_0^1 x^{j-1}x^{k-1}\,dx = \frac{1}{j+k-1}$$

for all $j, k = 1, \ldots, n$. Consider $v = \sum_{k=1}^{n} \alpha_k e_k \in V$ for $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$, and define the polynomial $p(x) = \sum_{k=1}^{n} \alpha_k x^{k-1}$. Note that $\overline{p(x)} = \sum_{k=1}^{n} \overline{\alpha_k} x^{k-1}$ since x is real. Then

$$\begin{aligned}
\langle Tv, v \rangle &= \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_k \overline{\alpha_j} \langle Te_k, e_j \rangle \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_k \overline{\alpha_j} \frac{1}{j + k - 1} \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_k \overline{\alpha_j} \int_0^1 x^{j-1} x^{k-1} \, dx \\
&= \int_0^1 \sum_{j=1}^{n} \sum_{k=1}^{n} \alpha_k \overline{\alpha_j} x^{j-1} x^{k-1} \, dx \\
&= \int_0^1 \sum_{k=1}^{n} \alpha_k x^{k-1} \sum_{j=1}^{n} \overline{\alpha_j} x^{j-1} \, dx \\
&= \int_0^1 p(x) \overline{p(x)} \, dx = \int_0^1 |p(x)|^2 \, dx \geq 0.
\end{aligned}$$

Hence, T is positive. Moreover, suppose $Tv = 0$. Then $\langle Tv, v \rangle = 0$, which implies $\int_0^1 |p(x)|^2 \, dx = 0$. Note that $|p(x)|^2$ is a continuous and nonnegative function on $[0, 1]$, so $\int_0^1 |p(x)|^2 \, dx = 0$ implies that $|p(x)|^2 = 0$ for all $x \in [0, 1]$. Then p has infinitely many zeros in $[0, 1]$ while a nonzero polynomial of degree at most $n - 1$ can have at most $n - 1$ zeros. Therefore, p is the zero polynomial. Hence, $\alpha_1 = \cdots = \alpha_n = 0$, so $v = 0$. Since V is finite-dimensional, it follows that T is invertible by 3.65. □

Solution 7C.22 Since $u \in V$, $V \neq \{0\}$. By 7.38(c), there exists an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T with nonnegative eigenvalues $\lambda_1, \ldots, \lambda_n$. Define

$$\lambda = \max_{1 \leq k \leq n} \lambda_k.$$

For any $v \in V$ with $\|v\| = 1$, write $v = \sum_{k=1}^{n} \alpha_k e_k$. Then

$$\|Tv\|^2 = \left\| \sum_{k=1}^{n} \alpha_k \lambda_k e_k \right\|^2 = \sum_{k=1}^{n} |\alpha_k|^2 \lambda_k^2 \leq \lambda^2 \sum_{k=1}^{n} |\alpha_k|^2 = \lambda^2 \|v\|^2 = \lambda^2.$$

Since $\|Tu\| \geq \|Tv\|$ for all $v \in V$ with $\|v\| = 1$, the above shows that $\|Tu\| \leq \lambda$. Moreover, if e_k is an eigenvector corresponding to the eigenvalue λ, then $\|Te_k\| = \lambda$. By hypothesis, $\|Tu\| \geq \|Te_k\| = \lambda$, so $\|Tu\| = \lambda$. Let $u = \sum_{k=1}^{n} \beta_k e_k$. Then $\sum_{k=1}^{n} |\beta_k|^2 = 1$, so

$$\|Tu\|^2 = \sum_{k=1}^{n} |\beta_k|^2 \lambda_k^2 = \lambda^2, \quad \text{and} \quad \lambda^2 \sum_{k=1}^{n} |\beta_k|^2 = \lambda^2.$$

Observe that

$$\sum_{k=1}^{n} |\beta_k|^2 (\lambda^2 - \lambda_k^2) = 0.$$

Since $\lambda_k \leq \lambda$ for all k, equality holds if and only if $|\beta_k|^2 = 0$ for all k with $\lambda_k < \lambda$. Hence, u is a linear combination of eigenvectors corresponding to the eigenvalue λ. Therefore, u is an eigenvector of T corresponding to the largest eigenvalue λ. □

Solution 7C.23

(a) $\Rightarrow$ Suppose $\langle \cdot, \cdot \rangle_T$ is an inner product on V. By positivity, $\langle Tv, v \rangle \geq 0$ for all $v \in V$. By conjugate symmetry,

$$\langle Tu, v \rangle = \langle u, v \rangle_T = \overline{\langle v, u \rangle_T} = \overline{\langle Tv, u \rangle} = \langle u, Tv \rangle$$

for all $u, v \in V$, so T is self-adjoint. Therefore, T is positive. If $Tv = 0$ for some $v \in V$, then $\langle Tv, v \rangle = 0$, which implies $v = 0$ by definiteness of the inner product $\langle \cdot, \cdot \rangle_T$. Hence, T is invertible by 3.65.

$\Leftarrow$ Suppose T is an invertible positive operator. By 7.38(f), there exists $R \in \text{End}(V)$ such that $R^*R = T$, so

$$\langle u, v \rangle_T = \langle Tu, v \rangle = \langle R^*Ru, v \rangle = \langle Ru, Rv \rangle$$

for all $u, v \in V$. Since $\langle \cdot, \cdot \rangle$ is an inner product and R is linear, $\langle \cdot, \cdot \rangle_T$ satisfies all the inner product axioms, except possibly definiteness. If $\langle v, v \rangle_T = 0$, then $\|Rv\|^2 = 0$, which implies $Rv = 0$ and hence $Tv = R^*(Rv) = 0$. Since T is invertible, this implies $v = 0$. Conversely, $\langle 0, 0 \rangle_T = \langle T0, 0 \rangle = \langle 0, 0 \rangle = 0$. Therefore, $\langle \cdot, \cdot \rangle_T$ is an inner product on V.

(b) Let $\langle \cdot, \cdot \rangle'$ be an inner product on V. For each $v \in V$, define $\phi_v : V \to \mathbb{F}$ by $\phi_v(u) = \langle u, v \rangle'$. Since $\langle \cdot, \cdot \rangle'$ is an inner product on V, ϕ_v is a linear functional on V. By the Riesz representation theorem, there exists a unique $Sv \in V$ such that $\phi_v(u) = \langle u, Sv \rangle$ for all $u \in V$. This defines a function $S : V \to V$. We claim that S is linear. To see this, let $v, w \in V$ and $\mu \in \mathbb{F}$. Then for all $u \in V$,

$$\langle u, S(v+w) \rangle = \langle u, v+w \rangle' = \langle u, v \rangle' + \langle u, w \rangle' = \langle u, Sv \rangle + \langle u, Sw \rangle = \langle u, Sv + Sw \rangle,$$

which implies $S(v + w) = Sv + Sw$ by the uniqueness part of the Riesz representation theorem. Similarly,

$$\langle u, S(\mu v) \rangle = \langle u, \mu v \rangle' = \overline{\mu} \langle u, v \rangle' = \overline{\mu} \langle u, Sv \rangle = \langle u, \mu Sv \rangle,$$

which implies $S(\mu v) = \mu Sv$ by the uniqueness part of the Riesz representation theorem. Hence, S is linear. By conjugate symmetry of both $\langle \cdot, \cdot \rangle'$ and $\langle \cdot, \cdot \rangle$,

$$\langle u, v \rangle' = \overline{\langle v, u \rangle'} = \overline{\langle v, Su \rangle} = \langle Su, v \rangle = \langle u, v \rangle_S$$

for all $u, v \in V$, so $\langle \cdot, \cdot \rangle' = \langle \cdot, \cdot \rangle_S$. Since $\langle \cdot, \cdot \rangle_S$ is an inner product, part (a) implies that S is positive and invertible.

□

Solution 7C.24 $\subseteq$ Let $v \in \ker(S+T)$. Then $(S+T)v = 0$, so

$$0 = \langle (S+T)v, v\rangle = \langle Sv, v\rangle + \langle Tv, v\rangle.$$

Since S and T are positive, $\langle Sv, v\rangle$ and $\langle Tv, v\rangle$ are both nonnegative. Because their sum is zero, we must have $\langle Sv, v\rangle = 0$ and $\langle Tv, v\rangle = 0$. By 7.43, $Sv = 0$ and $Tv = 0$, so $v \in \ker S \cap \ker T$.
$\supseteq$ Conversely, if $v \in \ker S \cap \ker T$, then $Sv = 0$ and $Tv = 0$, so $(S+T)v = Sv + Tv = 0$. Hence, $v \in \ker(S+T)$. □

Solution 7C.25 Let $D \in \text{End}(V)$ by $Df = f'$. By Exercise 7A31(a), $D^* = -D$ for $Df = f'$. Since $Tf = f''$, we have $T = D^2$. Therefore, $-T = -D^2 = (-D)(D) = D^*D$, which is positive by 7.38(f). □

§7D

Solution 7D.1 $\Rightarrow$ Let S be an isometry, and let e_1, e_2 be an orthonormal list of length two in V. By 7.49(c), $\langle Se_i, Se_j\rangle = \langle e_i, e_j\rangle$ for all $i, j \in \{1, 2\}$, so Se_1, Se_2 is an orthonormal list in W.
$\Leftarrow$ Let $u \in V$ with $\|u\| = 1$. By 6.36 and $\dim V \geq 2$, extend u to an orthonormal list u, v of length two in V. By hypothesis, Su, Sv is an orthonormal list in W; in particular, $\|Su\| = 1$. Now let $x \in V$. If $x = 0$, then $\|Sx\| = 0 = \|x\|$. Otherwise, $\frac{x}{\|x\|}$ has norm 1. Then

$$\|Sx\| = \left\| S\left(\frac{x}{\|x\|}\right)\right\| \|x\| = \|x\|.$$

Hence, S preserves norms so that S is an isometry. □

Solution 7D.2

Axler Note. *The phrase "T preserves orthogonality" means that $\langle Tu, Tv\rangle = 0$ for all $u, v \in V$ such that $\langle u, v\rangle = 0$.*

$\Rightarrow$ Let $T = \lambda S$, where $\lambda \in \mathbb{F}$ and $S \in \text{Hom}(V, W)$ is an isometry. Then for any $u, v \in V$ such that $\langle u, v\rangle = 0$, we have

$$\langle Tu, Tv\rangle = \langle \lambda Su, \lambda Sv\rangle = \lambda\overline{\lambda}\langle Su, Sv\rangle = |\lambda|^2\langle Su, Sv\rangle = |\lambda|^2\langle u, v\rangle = 0$$

by 7.49(c). Hence, T preserves orthogonality.

$\Leftarrow$ Since $T \neq 0$, $V \neq \{0\}$. Let $e_1, \ldots, e_n$ be an orthonormal basis of V, where $n = \dim V \geq 1$. For $j, k \in \{1, \ldots, n\}$ with $j \neq k$, we have $\langle e_j, e_k\rangle = 0$, so by assumption $\langle Te_j, Te_k\rangle = 0$. Since $\|e_j\| = \|e_k\|$ and $\langle e_k, e_j\rangle = \langle e_j, e_k\rangle = 0$, it follows that $\langle e_j + e_k, e_j - e_k\rangle = 0$. Therefore,

$$0 = \langle T(e_j+e_k), T(e_j-e_k)\rangle = \langle Te_j+Te_k, Te_j-Te_k\rangle = \langle Te_j, Te_j\rangle - \langle Te_k, Te_k\rangle = \|Te_j\|^2 - \|Te_k\|^2,$$

where $\langle Te_j, Te_k\rangle = \langle Te_k, Te_j\rangle = 0$ for $j \neq k$. Hence, $\|Te_j\| = \|Te_k\|$ for all $j, k \in \{1, \ldots, n\}$. Let $\lambda = \|Te_1\|$. Then $\|Te_j\| = \lambda$ for all $j = 1, \ldots, n$. If $\lambda = 0$, then $Te_j = 0$ for all j, so $T = 0$ on a basis. This implies $T = 0$, which contradicts

the assumption that $T \neq 0$. Hence, $\lambda > 0$. Define $S = \frac{1}{\lambda}T$. Then $\|Se_j\| = 1$ for all $j = 1, \ldots, n$. Moreover,

$$\langle Se_j, Se_k \rangle = \left\langle \frac{1}{\lambda}Te_j, \frac{1}{\lambda}Te_k \right\rangle = \frac{1}{\lambda^2}\langle Te_j, Te_k \rangle = 0$$

for $j \neq k$. Hence, $Se_1, \ldots, Se_n$ is an orthonormal list in W, and by 7.49(d), S is an isometry. Therefore, $T = \lambda S$ is a scalar multiple of an isometry. □

Solution 7D.3

(a) Let T_1 and T_2 be unitary operators on V. By 7.5(d) and 7.53(b), we have

$$(T_1T_2)^*(T_1T_2) = T_2^*T_1^*T_1T_2 = T_2^*T_2 = I, \quad (T_1T_2)(T_1T_2)^* = T_1T_2T_2^*T_1^* = T_1IT_1^* = T_1T_1^* = I.$$

Hence, T_1T_2 is a unitary operator by 7.53(b).

(b) Let T be a unitary operator on V. By 7.53(c), $T^{-1} = T^*$, and T^* is unitary by 7.53(f).

□

Solution 7D.4 Let $T = A + iB$. By 7.5(a), (b), and the fact that A and B are self-adjoint, we have $T^* = A - iB$. Therefore,

$$TT^* = A^2 + B^2 + i(BA - AB), \quad T^*T = A^2 + B^2 + i(AB - BA).$$

$\Rightarrow$ Suppose T is unitary. By 7.53(b), $TT^* = T^*T = I$. Then

$$TT^* - T^*T = i(BA - AB) - i(AB - BA) = 2i(BA - AB) = 0,$$

so $AB = BA$. Moreover, $TT^* = I$ implies $A^2 + B^2 + i(BA - AB) = I$, and since $AB = BA$, this simplifies to $A^2 + B^2 = I$.
$\Leftarrow$ Suppose $AB = BA$ and $A^2 + B^2 = I$. Substituting these into either TT^* or T^*T, we get

$$TT^* = T^*T = A^2 + B^2 = I.$$

Hence, T is unitary by 7.53(b). □

Solution 7D.5

(a) $\Rightarrow$ (b) Suppose S is a self-adjoint unitary operator. Then $S = S^*$. By 7.53(b), $S^*S = SS^* = I$, so $S^2 = I$. Let $P = \frac{1}{2}(S + I)$. By 7.5(a), (b), and (e), P is self-adjoint with

$$P^2 = \left(\frac{1}{2}(S + I)\right)^2 = \frac{1}{4}(S^2 + 2S + I) = \frac{1}{4}(I + 2S + I) = \frac{1}{2}(S + I) = P.$$

Since P is self-adjoint and $P^2 = P$, Exercise 7A20 gives that P is an orthogonal projection. Then $S = 2P - I$.

(b) $\Rightarrow$ (c) Suppose $S = 2P - I$ for some orthogonal projection P onto a subspace U of V. By 6.57, $Pu = u$ for $u \in U$, and $Pw = 0$ for $w \in U^\perp$. Hence,

$$Su = (2P - I)u = 2Pu - u = 2u - u = u, \quad Sw = (2P - I)w = 2Pw - w = 0 - w = -w$$

for $u \in U$ and $w \in U^\perp$.

(c) $\Rightarrow$ (a) Suppose there exists a subspace U of V such that $Su = u$ for every $u \in U$ and $Sw = -w$ for every $w \in U^\perp$. Since V is finite-dimensional, so is U. Let P be the orthogonal projection onto U. For any $v \in V$, write $v = u + w$ with $u \in U$ and $w \in U^\perp$ by 6.49. By 6.57(b) and (c),

$$(2P-I)v = (2P-I)(u+w) = 2P(u+w)-(u+w) = 2u-(u+w) = u-w = Su+Sw = Sv.$$

Hence, $S = 2P - I$. By Exercise 7A20, P is self-adjoint. It follows that $S^* = 2P^* - I = 2P - I = S$ by 7.5. By 6.57, $P^2 = P$, so

$$S^2 = (2P - I)^2 = 4P^2 - 4P + I = 4P - 4P + I = I$$

Therefore, $SS^* = S^*S = S^2 = I$, and S is a self-adjoint unitary operator by 7.53(b). □

Solution 7D.6 Let e_2, e_5, e_7 and f_2, f_5, f_7 be unit eigenvectors of T_1 and T_2 corresponding to the eigenvalues $2, 5, 7$, respectively. By 7.22, each list is an orthonormal list in $\mathbb{F}^3$. By 6.28, each list is an orthonormal basis of $\mathbb{F}^3$. By the linear map lemma, there exists a linear operator $S \in \text{End}(\mathbb{F}^3)$ such that $Se_2 = f_2$, $Se_5 = f_5$, and $Se_7 = f_7$. By 7.53(d), S is unitary, and $S^* = S^{-1}$ by 7.53(c). For each $k \in \{2, 5, 7\}$, we have $T_1 e_k = ke_k$ and $T_2 f_k = kf_k$, so

$$S^*T_2Se_k = S^*T_2f_k = S^*(kf_k) = kS^*f_k = kS^{-1}f_k = ke_k = T_1e_k.$$

Since T_1 and S^*T_2S agree on an orthonormal basis of $\mathbb{F}^3$, we have $T_1 = S^*T_2S$ by the linear map lemma. □

Solution 7D.7 Let e_1, e_2, e_3, e_4 be the standard basis of $\mathbb{F}^4$, which is an orthonormal basis. Consider the operators $T_1, T_2 \in \text{End}(\mathbb{F}^4)$ whose matrix representations with respect to e_1, e_2, e_3, e_4 are diagonal with diagonal entries $2, 2, 5, 7$ and $2, 5, 5, 7$, respectively. By 7.9, both operators are self-adjoint since real diagonal matrices equal their conjugate transposes. By 5.41, the eigenvalues of T_1 and T_2 are exactly $2, 5, 7$. Note that $E(2, T_1) = \ker(T_1 - 2I) = \text{span}(e_1, e_2)$ and $E(2, T_2) = \ker(T_2 - 2I) = \text{span}(e_1)$. Now suppose $S \in \text{End}(\mathbb{F}^4)$ is a unitary operator such that $T_1 = S^*T_2S$. By 7.53(c), $S^* = S^{-1}$. Then $ST_1 = S(S^*T_2S) = (SS^{-1})T_2S = T_2S$. For any $v \in E(2, T_1)$, we have $T_1 v = 2v$. Thus,

$$T_2(Sv) = S(T_1v) = S(2v) = 2Sv,$$

which shows that S maps $E(2, T_1)$ into $E(2, T_2)$. Moreover, $\dim E(2, T_1) = 2$ and $\dim E(2, T_2) = 1$. By 3.22, $S|_{E(2,T_1)}$ cannot be injective, and therefore S cannot be injective. Since a unitary operator must be invertible by 7.53(c), this is a contradiction. Hence, no such unitary operator exists. □

Solution 7D.8 The statement is false in general. Consider $V = \mathbb{F}^2$ with the standard orthonormal basis e_1, e_2. Define $S \in \text{End}(V)$ by $Se_1 = Se_2 = e_1$. Then $\|Se_1\| = \|Se_2\| = 1$, but S is not unitary since $S(e_1 - e_2) = 0$, which shows that S is not injective and therefore not invertible. □

Solution 7D.9 If $V = \{0\}$, then T is trivially unitary. Otherwise, $V \neq \{0\}$, so by Schur's theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $\text{Mat}(T)$ is upper triangular with respect to this basis; call this matrix A. Then for each $k = 1, \ldots, n$, $Te_k = \sum_{j=1}^{k} A_{j,k} e_j$. By 5.41, the diagonal entries $A_{k,k}$ are the eigenvalues of T, so $|A_{k,k}| = 1$ for each $k = 1, \ldots, n$. By 6.24 and the hypothesis,

$$1 \geq \|Te_k\|^2 = \sum_{j=1}^{k} |A_{j,k}|^2 = |A_{k,k}|^2 + \sum_{j=1}^{k-1} |A_{j,k}|^2 = 1 + \sum_{j=1}^{k-1} |A_{j,k}|^2,$$

which implies $0 \geq \sum_{j=1}^{k-1} |A_{j,k}|^2$. Then $A_{j,k} = 0$ for all $j < k$, so A is diagonal, and $Te_k = A_{k,k} e_k$ for each $k = 1, \ldots, n$. It follows that $e_1, \ldots, e_n$ is an orthonormal basis of eigenvectors of T with eigenvalues of absolute value 1. By 7.55, T is unitary. □

Solution 7D.10

(a) By 7.5, $I - T^2$ is self-adjoint. Then for any $v \in V$,

$$\langle (I - T^2)v, v \rangle = \langle v, v \rangle - \langle T^2 v, v \rangle = \langle v, v \rangle - \langle Tv, T^* v \rangle = \langle v, v \rangle - \langle Tv, Tv \rangle = \|v\|^2 - \|Tv\|^2 \geq 0,$$

where the final inequality follows from $\|Tv\| \leq \|v\|$ for all $v \in V$. Hence, $I - T^2$ is a positive operator.

(b) Let $B = \sqrt{I - T^2}$. By part (a) and 7.39, B is a positive self-adjoint operator. By Exercise 7C17, there exists a polynomial p such that $B = p(I - T^2)$, where p has real coefficients. Since T commutes with $I - T^2$, it commutes with B. Moreover,

$$T^2 + B^2 = T^2 + (I - T^2) = I.$$

By Exercise 7D4 ⇐ with $A = T$ and B as above, $T + iB$ is a unitary operator.

□

Solution 7D.11 Consider the set $\mathcal{B} = \{v \in V : \|v\| \leq 1\}$. For any $T \in \text{End}(V)$, let $T(\mathcal{B}) = \{Tv : v \in \mathcal{B}\}$. Now, if T is such that $T(\mathcal{B}) \subseteq \mathcal{B}$, then $\|Tv\| \leq 1$ for all $v \in \mathcal{B}$. If $v = 0$, then $\|T0\| = 0 = \|0\|$. If $v \neq 0$, then

$$\frac{\|Tv\|}{\|v\|} = \left\| T\left(\frac{v}{\|v\|}\right) \right\| \leq 1.$$

Hence, $\|Tv\| \leq \|v\|$ for all $v \in V$.

⇒ Suppose S is a unitary operator. Then S is also an isometry, so $S(\mathcal{B}) \subseteq \mathcal{B}$. By Exercise 7D3(b), S^{-1} is also unitary. Then for any $u \in \mathcal{B}$, $S^{-1}u \in \mathcal{B}$. Therefore, $u = S(S^{-1}u) \in S(\mathcal{B})$, so $S(\mathcal{B}) = \mathcal{B}$.

⇐ Now suppose $S(\mathcal{B}) = \mathcal{B}$. Let $w \in V$, and write $w = \|w\| w_0$ for some $w_0 \in \mathcal{B}$. Since $\mathcal{B} = S(\mathcal{B})$, there exists $v_0 \in \mathcal{B}$ such that $w_0 = Sv_0$. Hence, $w = \|w\| Sv_0 =$

$S(\|w\|v_0)$, so $w \in \operatorname{im} S$. Therefore, S is surjective, and hence invertible by 3.65. Also, if $u \in \mathcal{B}$, then $u = Sv$ for some $v \in \mathcal{B}$. Hence, $v = S^{-1}u \in \mathcal{B}$, so $S^{-1}(\mathcal{B}) \subseteq \mathcal{B}$. Applying the preliminary result from the beginning of the solution to S and S^{-1}, we get $\|Sx\| \leq \|x\|$ and $\|x\| = \|S^{-1}(Sx)\| \leq \|Sx\|$, so $\|Sx\| = \|x\|$ for all $x \in V$. Therefore, S is an isometry, and since it is also invertible, it is unitary. □

Solution 7D.12

Manual Remark. *The solution below uses 2 and $\frac{1}{2}$, but replacing them with λ and $\frac{1}{\lambda}$ for any nonzero $\lambda \in \mathbb{F}$ with $|\lambda| \neq 1$ gives a similar counterexample.*

The statement is false. Consider $\mathbb{F}^2$ with the Euclidean inner product, and let $S \in \operatorname{End}(\mathbb{F}^2)$ be defined by $S(x, y) = (2y, \frac{1}{2}x)$. Then

$$S^2(x, y) = S\left(2y, \frac{1}{2}x\right) = \left(2 \cdot \frac{1}{2}x, \frac{1}{2} \cdot 2y\right) = (x, y).$$

Hence, S is invertible with $S^{-1} = S$. Then $\|S^{-1}(x, y)\| = \|S(x, y)\|$ for all $(x, y) \in \mathbb{F}^2$. However, $\|S(0, 1)\| = 2 \neq 1 = \|(0, 1)\|$. Therefore, S is not unitary. □

Solution 7D.13 Let Q be an n-by-n matrix with complex entries. Define $T \in \operatorname{End}(\mathbb{C}^n)$ by $Tv = Qv$ for all $v \in \mathbb{C}^n$. Moreover, let $\mathbb{C}^n$ have the Euclidean inner product. If $e_1, \ldots, e_n$ denotes the standard (orthonormal) basis of $\mathbb{C}^n$, then Te_k is the kth column of Q. Hence, $\operatorname{Mat}(T) = Q$ with respect to $e_1, \ldots, e_n$. Then

$$\begin{aligned}
Q \text{ has orthonormal columns} &\iff Te_1, \ldots, Te_n \text{ is an orthonormal list} \\
&\iff Te_1, \ldots, Te_n \text{ is an orthonormal basis by 6.28} \\
&\iff T \text{ is unitary by 7.53(d)} \\
&\iff \text{the rows of } \operatorname{Mat}(T) = Q \text{ form an orthonormal basis by 7.53(e)} \\
&\iff \text{the rows of } Q \text{ form an orthonormal list by 6.28. } \square
\end{aligned}$$

Solution 7D.14 $\Rightarrow$ Note that $\|Sv\| = \|v\| = 1$ since S is unitary. By the Cauchy–Schwarz inequality, $|b| = |\langle Sv, v\rangle| \leq \|Sv\|\,\|v\| = 1$.
$\Leftarrow$ Suppose $|b| \leq 1$. By 6.36, we may extend v to an orthonormal basis $v, u, e_3, \ldots, e_n$ of V, which has length $\dim V \geq 2$. Let $\beta = \sqrt{1 - |b|^2}$. By the linear map lemma, there exists an operator $S \in \operatorname{End}(V)$ such that

$$Sv = bv + \beta u, \quad Su = -\beta v + \overline{b}u, \quad Se_k = e_k \text{ for } k \geq 3.$$

Then

$$\|Sv\|^2 = \|bv + \beta u\|^2 = |b|^2\|v\|^2 + |\beta|^2\|u\|^2 = |b|^2 + (1 - |b|^2) = 1,$$

and

$$\|Su\|^2 = \|-\beta v + \overline{b}u\|^2 = |\beta|^2\|v\|^2 + |b|^2\|u\|^2 = (1 - |b|^2) + |b|^2 = 1.$$

Moreover, noting that $\overline{\overline{b}} = b$, we have

$$\langle Sv, Su\rangle = \langle bv+\beta u, -\beta v+\overline{b}u\rangle = b(-\beta)\langle v, v\rangle + b^2\langle v, u\rangle - \beta^2\langle u, v\rangle + \beta b\langle u, u\rangle = -b\beta + \beta b = 0.$$

Lastly, $Sv, Su \in \operatorname{span}(v, u)$, which is orthogonal to all e_k for $k \geq 3$. Then the list $Sv, Su, Se_3, \ldots, Se_n$ is an orthonormal basis of V by 6.28, so S is unitary by 7.53(d). Finally,

$$\langle Sv, v\rangle = \langle bv + \beta u, v\rangle = b\langle v, v\rangle + \beta\langle u, v\rangle = b. \qquad \square$$

Solution 7D.15

Axler Note. *The function $z \mapsto i(z+1)(z-1)^{-1}$ maps the unit circle in $\mathbb{C}$ (except for the point 1) to $\mathbb{R}$. Thus (b) illustrates the analogy between the unitary operators and the unit circle in $\mathbb{C}$, along with the analogy between the self-adjoint operators and $\mathbb{R}$.*

Define $A = (T + I)(T - I)^{-1}$. Observe that $T + I = (T - I) + 2I$. Then $A = I + 2(T - I)^{-1}$.

(a) By 7.5, $A^* = I + 2(T^* - I)^{-1}$. By 7.53(c), $T^* = T^{-1}$, so $T^* - I = -T^{-1}(T - I)$, and

$$(T^* - I)^{-1} = -(T - I)^{-1}T = -(T - I)^{-1}((T - I) + I) = -I - (T - I)^{-1}.$$

Therefore, $A^* = -I - 2(T - I)^{-1} = -A$.

(b) Using 7.5(b) and part (a), $(iA)^* = \bar{\imath}A^* = (-i)(-A) = iA$. $\square$

Solution 7D.16 Observe that every eigenvalue of T is real by 7.12. Therefore, neither i nor $-i$ is an eigenvalue of T so that $T - iI$ and $T + iI$ are invertible. Define $S = (T + iI)(T - iI)^{-1}$. By 7.5, $(T + iI)^* = T^* - iI = T - iI$ and $((T - iI)^{-1})^* = ((T - iI)^*)^{-1} = (T + iI)^{-1}$. Thus,

$$S^* = (T + iI)^{-1}(T - iI).$$

Because $T - iI$ and $T + iI$ commute,

$$S^*S = (T + iI)^{-1}(T - iI)(T + iI)(T - iI)^{-1} = I,$$

showing that S is an isometry on V by 7.49(b), whence S is unitary.
Now, suppose $Sx = x$ for some $x \in V$. By definition of S, $(T + iI)(T - iI)^{-1}x = x$, so $(T + iI)y = x$ where $y = (T - iI)^{-1}x$. Since $x = (T - iI)y$, we have $(T + iI)y = (T - iI)y$, so $2iy = 0$, and hence $y = 0$. Therefore, $x = (T - iI)y = 0$, showing that 1 is not an eigenvalue of S. $\square$

Solution 7D.17 Let Q be an n-by-n matrix. Define $T \in \operatorname{End}(\mathbb{F}^n)$ by $Tv = Qv$ for all $v \in \mathbb{F}^n$, so $\operatorname{Mat}(T) = Q$ with respect to the standard basis of $\mathbb{F}^n$, which is orthonormal with respect to the Euclidean inner product.
(a) $\Leftrightarrow$ (c) Since $Tv = Qv$ for all $v \in \mathbb{F}^n$, (c) states precisely that T is an isometry. By the equivalence of 7.49(a) and (e), T is an isometry if and only if the columns of $\operatorname{Mat}(T) = Q$ are orthonormal, which is (a) by definition.
(a) $\Leftrightarrow Q^*Q = I$, (b) $\Leftrightarrow QQ^* = I$ For $j, k \in \{1, \ldots, n\}$, consider the j, kth entry of both Q^*Q and QQ^*. By 3.46 and the definition of the conjugate transpose, we have

$$(Q^*Q)_{j,k} = \sum_{\ell=1}^{n} \overline{Q_{\ell,j}}\, Q_{\ell,k} = \langle Q_{\cdot,k}, Q_{\cdot,j}\rangle, \quad (QQ^*)_{j,k} = \sum_{\ell=1}^{n} Q_{j,\ell}\overline{Q_{k,\ell}} = \langle Q_{j,\cdot}, Q_{k,\cdot}\rangle.$$

Note that $Q_{j,\cdot}$ and $Q_{k,\cdot}$ are treated as vectors in $\mathbb{F}^n$. Therefore, $Q^*Q = I$ if and only if the columns of Q are orthonormal, which is (a), and $QQ^* = I$ if and only if the rows of Q are orthonormal, which is (b).

$\boxed{Q^*Q = I \Leftrightarrow QQ^* = I}$ By 7.9 and 3.43, $\text{Mat}(T^*T) = Q^*Q$ and $\text{Mat}(TT^*) = QQ^*$. Then $Q^*Q = I$ if and only if $T^*T = I$, and $QQ^* = I$ if and only if $TT^* = I$. By 3.68, $T^*T = I$ if and only if $TT^* = I$. Therefore, $Q^*Q = I$ if and only if $QQ^* = I$, so each condition is equivalent to (d).

This shows that (a) is equivalent to (c) and that (a) is equivalent to (d), which is equivalent to (b). □

Solution 7D.18 Let A be a symmetric n-by-n matrix with real entries. Define $T \in \text{End}(\mathbb{R}^n)$ by $Tv = Av$ for all $v \in \mathbb{R}^n$, so $\text{Mat}(T) = A$ with respect to the standard orthonormal basis of $\mathbb{R}^n$. Since A is real and symmetric, $A^* = A^t = A$. Then $\text{Mat}(T^*) = A^* = A = \text{Mat}(T)$, so T is self-adjoint by 7.9. By the real spectral theorem, there exists an orthonormal basis $f_1, \ldots, f_n$ of $\mathbb{R}^n$ such that $Tf_k = \lambda_k f_k$ for some real numbers $\lambda_1, \ldots, \lambda_n$.

Define Q to be the matrix whose columns are the vectors $f_1, \ldots, f_n$. Moreover, let D be the diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$ on its diagonal. By definition, Q is a matrix with real entries and orthonormal columns, so it is unitary. Moreover, the kth column of AQ is $Af_k = Tf_k = \lambda_k f_k$, which is also the kth column of QD for each $k = 1, \ldots, n$. Hence, $AQ = QD$. By 7.57(d), $Q^*Q = I$, so $Q^*AQ = Q^*QD = D$. □

Solution 7D.19 Let $e_0, e_1, \ldots, e_{n-1}$ be the standard orthonormal basis of $\mathbb{C}^n$, so the jth component of $\mathcal{F}e_k$ is

$$\omega_j(e_k) = \frac{1}{\sqrt{n}} \sum_{p=0}^{n-1} (e_k)_p e^{-2\pi ijp/n} = \frac{1}{\sqrt{n}} e^{-2\pi ijk/n},$$

where $(e_k)_p = \delta_{pk}$.

(a) Let $m, k \in \{0, 1, \ldots, n-1\}$. Then

$$\langle \mathcal{F}e_m, \mathcal{F}e_k \rangle = \sum_{j=0}^{n-1} \omega_j(e_m)\overline{\omega_j(e_k)} = \sum_{j=0}^{n-1} \frac{1}{\sqrt{n}} e^{-2\pi ijm/n} . \frac{1}{\sqrt{n}} e^{2\pi ijk/n} = \frac{1}{n} \sum_{j=0}^{n-1} e^{2\pi ij(k-m)/n}.$$

If $m = k$, then $\langle \mathcal{F}e_m, \mathcal{F}e_k \rangle = 1$. Otherwise, let $\zeta = e^{2\pi i(k-m)/n}$. Then $\zeta^n = 1$ and $\zeta \neq 1$, since $0 < |k - m| < n$. Therefore,

$$\langle \mathcal{F}e_m, \mathcal{F}e_k \rangle = \frac{1}{n} \sum_{j=0}^{n-1} \zeta^j = \frac{1}{n} \cdot \frac{1 - \zeta^n}{1 - \zeta} = 0.$$

Hence, $\mathcal{F}e_0, \ldots, \mathcal{F}e_{n-1}$ is an orthonormal list in $\mathbb{C}^n$. By 6.28, $\mathcal{F}e_0, \ldots, \mathcal{F}e_{n-1}$ is an orthonormal basis of $\mathbb{C}^n$, so $\mathcal{F}$ is unitary on $\mathbb{C}^n$ by 7.53(d).

(b) Define $R \in \text{End}(\mathbb{C}^n)$ by $R(z_0, z_1, \ldots, z_{n-1}) = (z_{n-1}, z_{n-2}, \ldots, z_0)$. Then for any $z \in \mathbb{C}^n$ and $j \in \{0, \ldots, n-1\}$,

$$(\mathcal{F}^2 z)_j = \frac{1}{\sqrt{n}} \sum_{m=0}^{n-1} (\mathcal{F}z)_m e^{-2\pi ijm/n} = \frac{1}{n} \sum_{m=0}^{n-1} \sum_{k=0}^{n-1} z_k e^{-2\pi im(k+j)/n} = \sum_{k=0}^{n-1} z_k \left(\frac{1}{n} \sum_{m=0}^{n-1} e^{-2\pi im(k+j)/n} \right)$$

where $r_k = e^{-2\pi i(k+j)/n}$. Observe that $0 \le k + j \le 2n - 2$, so $r_k \ne 1$ unless $k + j = 0$ or $k + j = n$ in which case the inner sum evaluates to 1. Otherwise, we have $r_k \ne 1$ and $r_k^n = 1$, so

$$\frac{1}{n}\sum_{m=0}^{n-1} r_k^m = \frac{1}{n} \cdot \frac{1 - r_k^n}{1 - r_k} = 0.$$

Observe that for $k \in \{0, \dots, n-1\}$, one such k exists for which $k + j = 0$ or $k + j = n$, namely $k = 0$ if $j = 0$ or $k = n - j$ if $1 \le j \le n - 1$. Hence, $(\mathcal{F}^2 z)_0 = z_0$ and for $j \ge 1$, $(\mathcal{F}^2 z)_j = z_{n-j}$, whereupon $\mathcal{F}^2 z = Rz$. Since R fixes the 0th coordinate and reverses the jth and $(n - j)$th coordinates for $j \ge 1$, R is its own inverse, i.e., $R^2 = I$. From this, we get $\mathcal{F}(\mathcal{F}R) = \mathcal{F}^2 R = R^2 = I$. By 3.68, $(\mathcal{F}R)\mathcal{F} = I$, and hence $\mathcal{F}^{-1} = \mathcal{F}R$. Because $z_n = z_0$,

$$\mathcal{F}^{-1}(z_0, z_1, \dots, z_{n-1}) = (\mathcal{F}R)(z_0, z_1, \dots, z_{n-1}) = \mathcal{F}(z_0, z_{n-1}, \dots, z_1).$$

(c) By part (b), $\mathcal{F}^2 = R$ and $R^2 = I$, so $\mathcal{F}^4 = R^2 = I$. □

Solution 7D.20 Since A is n-by-n and has linearly independent columns, column rank $A = n$. By 3.57, row rank A = column rank $A = n$, so the rows of A are linearly independent as well. Since the columns of A^* are the conjugate transposes of the rows of A, the columns of A^* are linearly independent. By 7.58, there exist unique matrices Q_1 and R_1 such that Q_1 is unitary, R_1 is upper triangular with positive diagonal entries, and $A^* = Q_1 R_1$. Taking conjugate transposes gives $A = R_1^* Q_1^*$, where R_1^* is lower triangular with the same positive diagonal entries as R_1, and Q_1^* is unitary by 7.57(d). Let $R = R_1^*$ and $Q = Q_1^*$, so $A = RQ$ where R and Q have the desired properties.
To see uniqueness of R and Q, suppose $A = RQ = R_0 Q_0$ where R_0 is lower triangular with positive diagonal entries and Q_0 is unitary. Taking conjugate transposes again yields $A^* = Q^* R^* = Q_0^* R_0^*$, where R^* and R_0^* are upper triangular with positive diagonal entries and Q^* and Q_0^* are unitary. By the uniqueness part of 7.58, we must have $R^* = R_0^*$ and $Q^* = Q_0^*$, which implies $R = R_0$ and $Q = Q_0$. □

§7E

Solution 7E.1 Lemma 1. Suppose $T \in \mathrm{Hom}(V, W)$ and $v_1, \dots, v_n$ is a basis of V with $T^*T v_k = \mu_k v_k$ for $k = 1, \dots, n$. Then the singular values of T are $\sqrt{\mu_1}, \dots, \sqrt{\mu_n}$ sorted into decreasing order.
Proof. Each μ_k is an eigenvalue of T^*T since $v_k \ne 0$. By 7.64(a), T^*T is positive, so $\mu_k \ge 0$ for all $k = 1, \dots, n$ by 7.38(b). Since $T^*T v_k = \mu_k v_k$ for each k, the matrix $\mathrm{Mat}(T^*T, (v_1, \dots, v_n))$ is diagonal with diagonal entries $\mu_1, \dots, \mu_n$. By Exercise 5D2, each eigenvalue λ of T^*T appears in the list $\mu_1, \dots, \mu_n$ exactly $\dim E(\lambda, T^*T)$ times. By the definition of singular values, the singular values of T are precisely $\sqrt{\mu_1}, \dots, \sqrt{\mu_n}$ sorted into decreasing order. □

$\Rightarrow$ If $T = 0$, then $\operatorname{im} T = \{0\}$. Hence, $\dim \operatorname{im} T = 0$. By 7.68(b), T has no positive singular values. Since all singular values are nonnegative, every singular value of T is 0.

$\Leftarrow$ Suppose all singular values of T are 0. Then T has no positive singular values, so $\dim \operatorname{im} T = 0$ by 7.68(b). Therefore, $\operatorname{im} T = \{0\}$, so $T = 0$. □

Solution 7E.2

Axler Note. *The vectors v, w satisfying both equations above are called a* Schmidt pair. *Erhard Schmidt introduced the concept of singular values in 1907.*

$\Rightarrow$ Suppose s is a singular value of T. Since s is a square root of an eigenvalue of T^*T, s^2 is an eigenvalue of T^*T. Let $v \in V$ be nonzero such that $T^*Tv = s^2v$. Let $w = \frac{1}{s}Tv$. It follows that $Tv = sw$, and

$$T^*w = T^*\left(\frac{1}{s}Tv\right) = \frac{1}{s}T^*Tv = \frac{1}{s}s^2v = sv.$$

Since $sv \neq 0$, we have $w \neq 0$.

$\Leftarrow$ Suppose there exist nonzero vectors $v \in V$ and $w \in W$ such that $Tv = sw$ and $T^*w = sv$. Observe that $T^*Tv = sT^*w = s^2v$. Then s^2 is an eigenvalue of T^*T. Since $s > 0$, $\sqrt{s^2} = s$ is a singular value of T. □

Solution 7E.3 Consider $T \in \operatorname{End}(\mathbb{C}^2)$ defined by $T(z, w) = (5w, 0)$. With respect to the standard orthonormal basis e_1, e_2 of $\mathbb{C}^2$, the matrix of T is

$$\begin{pmatrix} 0 & 5 \\ 0 & 0 \end{pmatrix}.$$

Since the matrix of T is upper triangular with all diagonal entries 0, 0 is the only eigenvalue of T by 5.41. Using 7.9 and 3.43,

$$\operatorname{Mat}(T^*T) = \begin{pmatrix} 0 & 0 \\ 0 & 25 \end{pmatrix}.$$

Then $T^*Te_1 = 0$ and $T^*Te_2 = 25e_2$. Therefore, by Lemma 1, the singular values of T are $\sqrt{25} = 5$ and $\sqrt{0} = 0$. □

Solution 7E.4 Let $n = \dim V$. Let $\mathcal{S} = \{\|Tv\| : v \in V \text{ and } \|v\| = 1\}$. Since T has a largest and a smallest singular value, $V \neq \{0\}$, so $n \geq 1$. Since T^*T is positive by 7.64(a), 7.38(c) gives an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T^*T, say $T^*Te_k = \mu_k e_k$ for each $k = 1, \ldots, n$. After reordering, Lemma 1 shows that $\mu_k = s_k^2$ for each $k = 1, \ldots, n$.

$\subseteq$ Suppose $\|v\| = 1$. Let $c_k = \langle v, e_k \rangle$ for each $k = 1, \ldots, n$. By 6.30(a) and 6.30(b), $v = \sum_{k=1}^n c_k e_k$ and $\sum_{k=1}^n |c_k|^2 = \|v\|^2 = 1$. We then have

$$\|Tv\|^2 = \langle T^*Tv, v \rangle = \left\langle \sum_{k=1}^n c_k s_k^2 e_k, \sum_{j=1}^n c_j e_j \right\rangle = \sum_{k=1}^n s_k^2 |c_k|^2.$$

Therefore, $\|Tv\|^2$ is a weighted average of $s_1^2, \ldots, s_n^2$ with weights $|c_k|^2$, so $s_n^2 \leq \|Tv\|^2 \leq s_1^2$, and taking square roots gives $s_n \leq \|Tv\| \leq s_1$. Hence, $\mathcal{S} \subseteq [s_n, s_1]$.

$\supseteq$ Let $s \in [s_n, s_1]$. If $s_1 = s_n$, take $v = e_1$. Otherwise, $n \geq 2$ and $s_n < s_1$. Define

$$\theta = \frac{s^2 - s_n^2}{s_1^2 - s_n^2}, \quad v = \sqrt{\theta} e_1 + \sqrt{1-\theta} e_n.$$

Since $s_n < s_1$ and $s \in [s_n, s_1]$, we have $\theta \in [0, 1]$. By orthonormality of e_1 and e_n, we have $\|v\|^2 = \theta + (1-\theta) = 1$. Additionally, $\langle v, e_1 \rangle = \sqrt{\theta}$, $\langle v, e_n \rangle = \sqrt{1-\theta}$, and $\langle v, e_k \rangle = 0$ for $1 < k < n$. Thus,

$$\|Tv\|^2 = s_1^2 \theta + s_n^2 (1-\theta) = s^2.$$

Hence, $\|Tv\| = s$, so $s \in \mathcal{S}$. Therefore, $[s_n, s_1] \subseteq \mathcal{S}$. Combining with the previous inclusion gives $\mathcal{S} = [s_n, s_1]$. □

Solution 7E.5 With respect to the standard orthonormal basis e_1, e_2 of $\mathbb{C}^2$, the matrix of T is

$$\mathrm{Mat}(T) = \begin{pmatrix} 0 & -4 \\ 1 & 0 \end{pmatrix},$$

so by 7.9, the matrix of T^* with respect to the same basis is

$$\mathrm{Mat}(T^*) = \begin{pmatrix} 0 & 1 \\ -4 & 0 \end{pmatrix}.$$

By 3.43, $\mathrm{Mat}(T^*T)$ is diagonal with diagonal entries 1 and 16. Therefore, $T^*Te_1 = e_1$ and $T^*Te_2 = 16e_2$. By Lemma 1, the singular values of T are $\sqrt{16} = 4$ and $\sqrt{1} = 1$. □

Solution 7E.6 Let e_1, e_2, e_3 be the orthonormal basis of $\mathbb{R}[X]_{\leq 2}$ given by

$$e_1 = \sqrt{\frac{1}{2}}, \quad e_2 = \sqrt{\frac{3}{2}} x, \quad e_3 = \sqrt{\frac{45}{8}} \left(x^2 - \frac{1}{3} \right)$$

as described in Example 6.34. Then

$$De_1 = 0, \quad De_2 = \sqrt{\frac{3}{2}} = \sqrt{3} e_1, \quad De_3 = \sqrt{\frac{45}{8}} \cdot 2x = \sqrt{15} e_2.$$

Therefore, the matrix of D with respect to the orthonormal basis e_1, e_2, e_3 is

$$\mathrm{Mat}(D) = \begin{pmatrix} 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{15} \\ 0 & 0 & 0 \end{pmatrix}.$$

By 7.9, the matrix of D^* with respect to the same basis is

$$\mathrm{Mat}(D^*) = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{3} & 0 & 0 \\ 0 & \sqrt{15} & 0 \end{pmatrix}.$$

By 3.43, $\mathrm{Mat}(D^*D)$ is diagonal with diagonal entries 0, 3, and 15. Therefore, $D^*De_1 = 0$, $D^*De_2 = 3e_2$, and $D^*De_3 = 15e_3$. By Lemma 1, the singular values of D are $\sqrt{15}$, $\sqrt{3}$, and 0. □

Solution 7E.7 If T is self-adjoint, then $T^*T = T^2 = TT^*$, so T is normal. Hence, T is normal in either case. If $\mathbb{F} = \mathbb{R}$, then T is self-adjoint, so the real spectral theorem applies. If $\mathbb{F} = \mathbb{C}$ and T is normal, then the complex spectral theorem applies. In either case, we obtain an orthonormal basis $e_1, \ldots, e_{\dim V}$ of V with $Te_k = \mu_k e_k$ for $k = 1, \ldots, \dim V$, where $\mu_k \in \mathbb{F}$. Then $\operatorname{Mat}(T, (e_1, \ldots, e_{\dim V}))$ is diagonal with diagonal entries $\mu_1, \ldots, \mu_{\dim V}$. Since $e_k \neq 0$, each μ_k is an eigenvalue of T. By Exercise 5D2, each eigenvalue λ of T appears in the list exactly $\dim E(\lambda, T)$ times, so $\mu_1, \ldots, \mu_{\dim V}$ is a rearrangement of $\lambda_1, \ldots, \lambda_n$. In particular, $\dim V = n$. Thus, we may reorder the basis so that $\mu_k = \lambda_k$ for $k = 1, \ldots, n$. By 7.21(e), $T^*e_k = \overline{\lambda_k} e_k$ for $k = 1, \ldots, n$, so

$$T^*Te_k = \lambda_k T^* e_k = \lambda_k \overline{\lambda_k} e_k = |\lambda_k|^2 e_k.$$

By Lemma 1, the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$, sorted into decreasing order. □

Solution 7E.8 Consider the maps $S_k \in \operatorname{Hom}(V, W)$ defined by $S_k v = \langle v, e_k \rangle f_k$ for $k = 1, \ldots, m$. Then $T = \sum_{k=1}^m s_k S_k$. Following the calculations in Example 7.3, $S_k^* w = \langle w, f_k \rangle e_k$ for $k = 1, \ldots, m$ and $w \in W$. Since $s_k \in \mathbb{R}$, by 7.5(a) and (b), we have

$$T^* w = \sum_{k=1}^m s_k S_k^* w = \sum_{k=1}^m s_k \langle w, f_k \rangle e_k.$$

Since $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ are orthonormal lists, we have $Te_k = s_k f_k$ and $T^* f_k = s_k e_k$ for $k = 1, \ldots, m$. Hence, $T^*Te_k = T^*(s_k f_k) = s_k^2 e_k$ for $k = 1, \ldots, m$.

(a) The computation of Tv for $v \in V$ gives $\operatorname{im} T \subseteq \operatorname{span}(f_1, \ldots, f_m)$. Since $f_k = Te_k / s_k$, we have $f_k \in \operatorname{im} T$ for $k = 1, \ldots, m$. Hence, $\operatorname{span}(f_1, \ldots, f_m) \subseteq \operatorname{im} T$, and equality follows. By 6.25, orthonormal lists are linearly independent, so $f_1, \ldots, f_m$ is an orthonormal basis of $\operatorname{im} T$.

(b) Using the same argument as in part (a) with T^* in place of T, we get $\operatorname{im} T^* = \operatorname{span}(e_1, \ldots, e_m)$. Since $\operatorname{im} T^* = (\ker T)^\perp$ by 7.6(b), it follows that $e_1, \ldots, e_m$ is an orthonormal basis of $(\ker T)^\perp$.

(c) By 6.36, we may extend $e_1, \ldots, e_m$ to an orthonormal basis $e_1, \ldots, e_n$ of V. For $k \leq m$, the beginning calculation shows that $T^*Te_k = s_k^2 e_k$. For $k > m$, we have $Te_k = 0$, so $T^*Te_k = 0$. By Lemma 1 and the hypothesis that $s_1 \geq \cdots \geq s_m > 0$, it follows that the singular values of T are $s_1, \ldots, s_m$ together with $n - m$ zeros. Thus, the positive singular values of T are exactly $s_1, \ldots, s_m$.

(d) By the beginning calculation, $T^*Te_k = s_k^2 e_k$ and $e_k \neq 0$ since $\|e_k\| = 1$ for $k = 1, \ldots, m$. Hence, e_k is indeed an eigenvector of T^*T corresponding to the eigenvalue s_k^2.

(e) For any $w \in W$, we have

$$TT^* w = \sum_{k=1}^m s_k \langle w, f_k \rangle Te_k = \sum_{k=1}^m s_k \langle w, f_k \rangle s_k f_k = \sum_{k=1}^m s_k^2 \langle w, f_k \rangle f_k. \qquad \square$$

Solution 7E.9 If $T = 0$, then $T^* = 0$ by 7.5(b), so neither T nor T^* has any positive singular values, and the statement holds trivially. Now assume $T \neq 0$. Let $s_1, \ldots, s_m$ denote the positive singular values of T. By definition, $s_1 \geq \cdots \geq s_m > 0$. Moreover, $m = \dim \operatorname{im} T \geq 1$ by 7.68(b), since $T \neq 0$. By the singular value decomposition, there exist orthonormal lists $e_1, \ldots, e_m$ in V and $f_1, \ldots, f_m$ in W such that

$$Tv = \sum_{k=1}^{m} s_k \langle v, e_k \rangle f_k$$

for all $v \in V$. By 7.75,

$$T^* w = \sum_{k=1}^{m} s_k \langle w, f_k \rangle e_k$$

for all $w \in W$. Applying Exercise 7E8(c) to $T^* \in \operatorname{Hom}(W, V)$ with the orthonormal lists $f_1, \ldots, f_m$ in W and $e_1, \ldots, e_m$ in V, the positive singular values of T^* are exactly $s_1, \ldots, s_m$. Hence, T and T^* have the same positive singular values. □

Solution 7E.10 If $\dim V = 0$, then $\dim W = 0$ by 3.70, so neither T nor T^{-1} has any singular values, and the statement holds trivially. Now assume $\dim V > 0$. Since T is invertible, T is injective. By 7.68(a) and the definition of singular values, $s_1 \geq \cdots \geq s_n > 0$. Since all singular values are positive, the positive singular values of T are exactly $s_1, \ldots, s_n$. By the singular value decomposition, there exist orthonormal lists $e_1, \ldots, e_n$ in V and $f_1, \ldots, f_n$ in W such that

$$Tv = \sum_{k=1}^{n} s_k \langle v, e_k \rangle f_k$$

for all $v \in V$. By 6.69(a), $T^\dagger = T^{-1}$. Since $s_1 \geq \cdots \geq s_n > 0$, we have $\frac{1}{s_n} \geq \cdots \geq \frac{1}{s_1} > 0$. Reordering the sum, we have by 7.75 that

$$T^{-1} w = \sum_{k=1}^{n} \frac{1}{s_{n-k+1}} \langle w, f_{n-k+1} \rangle e_{n-k+1}$$

for all $w \in W$. Applying Exercise 7E8(c) to $T^{-1} \in \operatorname{Hom}(W, V)$ with the orthonormal lists $f_n, \ldots, f_1$ in W and $e_n, \ldots, e_1$ in V, the positive singular values of T^{-1} are exactly $\frac{1}{s_n}, \ldots, \frac{1}{s_1}$. By 3.70, $\dim V = \dim W = n$, so T^{-1} has exactly n singular values. Since T^{-1} is injective, 0 is not a singular value of T^{-1} by 7.68(a). Hence, the singular values of T^{-1} are exactly $\frac{1}{s_n}, \ldots, \frac{1}{s_1}$. □

Solution 7E.11

(a) Let m denote the number of positive singular values of T. By the singular value decomposition, there exist orthonormal lists $e_1, \ldots, e_m$ in V and $f_1, \ldots, f_m$ in W such that $Te_k = s_k f_k$ for $k = 1, \ldots, m$. By 6.36, extend $e_1, \ldots, e_m$ to an orthonormal basis $e_1, \ldots, e_n$ of V. By the formula in the singular value decomposition, $Te_k = 0$ for $k > m$ since $\langle e_k, e_j \rangle = 0$ for $j = 1, \ldots, m$. Moreover, $s_k = 0$ for $k > m$, since T has exactly n singular values. Therefore,

$$\sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{m} \|s_k f_k\|^2 = \sum_{k=1}^{m} s_k^2 = \sum_{k=1}^{n} s_k^2.$$

By Exercise 7A5, the left-hand side is the same for any orthonormal basis of V. Therefore, the equality holds for the given orthonormal basis $v_1, \ldots, v_n$ as well.

(b) Since $T \in \text{End}(V)$ is a positive operator, it is self-adjoint. By Exercise 7E7, the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of T, and each eigenvalue λ is repeated exactly $\dim E(\lambda, T)$ times. By 7.38(b), the eigenvalues of T are nonnegative, so $s_1, \ldots, s_n$ are exactly the eigenvalues of T. By 7.38(f), there exists an operator $R \in \text{End}(V)$ such that $T = R^*R$. By part (a), the sum of the squares of the norms of Rv_k equals the sum of the squares of the singular values of R. Since $R^*R = T$, the singular values of R are, by definition, $\sqrt{s_1}, \ldots, \sqrt{s_n}$. Hence,

$$\sum_{k=1}^{n} \langle Tv_k, v_k \rangle = \sum_{k=1}^{n} \langle R^*Rv_k, v_k \rangle = \sum_{k=1}^{n} \langle Rv_k, Rv_k \rangle = \sum_{k=1}^{n} \|Rv_k\|^2 = \sum_{k=1}^{n} s_k. \quad \square$$

Solution 7E.12

(a) Consider $T \in \text{End}(\mathbb{C}^2)$ defined by $T(z, w) = (5w, 0)$. By Exercise 7E3, its singular values are 5 and 0. However, $T^2(z, w) = T(5w, 0) = (0, 0)$, so $T^2 = 0$ and its singular values are 0 and 0, not 25 and 0.

(b) Using 7.5(c), (d), and normality of T,

$$(T^2)^*T^2 = T^*(T^*T)T = T^*(TT^*)T = (T^*T)^2 = (T^*T)^*(T^*T).$$

By definition, T^2 and T^*T have the same singular values. Since T^*T is a positive operator by 7.64(a), its eigenvalues are nonnegative by 7.38(b). By Exercise 7E7 applied to T^*T, its singular values are its eigenvalues λ, each repeated $\dim E(\lambda, T^*T)$ times. The eigenvalues of T^*T are the squares of the singular values of T, so the singular values of T^2 are the squares of the singular values of T.

$\square$

Solution 7E.13 Let $n = \dim V$, let $s_1 \geq \cdots \geq s_n \geq 0$ denote the singular values of T_2, and let m be the number of positive singular values of T_2. If $m = 0$, then T_2 has no positive singular values, so by the formula in the singular value decomposition, $T_2 = 0$. If T_1 and T_2 have the same singular values, the same reasoning yields $T_1 = 0$, so we may take $S_1 = S_2 = I$ and the result holds trivially. Conversely, if $T_1 = S_1T_2S_2$ for unitary $S_1, S_2 \in \text{End}(V)$, then $T_1 = S_1 0 S_2 = 0$, so $T_1^*T_1 = 0$. Since $E(0, T_1^*T_1) = V$, 0 is the only eigenvalue of $T_1^*T_1$, and $\dim E(0, T_1^*T_1) = n$. Therefore, the singular values of T_1 are all 0, which are the same as the singular values of T_2. Hence, we may assume $m > 0$ for the remainder of the proof.

$\Rightarrow$ Suppose T_1 and T_2 have the same singular values. By the singular value decomposition, there exist orthonormal lists $e_1, \ldots, e_m$, $f_1, \ldots, f_m$, $e'_1, \ldots, e'_m$, and $f'_1, \ldots, f'_m$ in V such that

$$T_1 v = \sum_{i=1}^{m} s_i \langle v, e_i \rangle f_i, \quad T_2 v = \sum_{i=1}^{m} s_i \langle v, e'_i \rangle f'_i$$

for every $v \in V$. By 6.36, extend all orthonormal lists $e_1, \ldots, e_m$, $f_1, \ldots, f_m$, $e'_1, \ldots, e'_m$, and $f'_1, \ldots, f'_m$ to orthonormal bases $e_1, \ldots, e_n$, $f_1, \ldots, f_n$, $e'_1, \ldots, e'_n$, and $f'_1, \ldots, f'_n$ of V. By the orthonormality of these lists, $T_1 e_i = s_i f_i$ and $T_2 e'_i = s_i f'_i$ for $i \leq m$, and $T_1 e_i = 0$ and $T_2 e'_i = 0$ for $i > m$. By the linear map lemma, we may define operators $S_1, S_2 \in \text{End}(V)$ such that $S_1 f'_i = f_i$ and $S_2 e_i = e'_i$ for $i = 1, \ldots, n$. By 7.53(d), S_1 and S_2 are unitary. Moreover, for $i = 1, \ldots, m$, we have

$$S_1 T_2 S_2 e_i = S_1 T_2 e'_i = S_1(s_i f'_i) = s_i f_i = T_1 e_i,$$

and $S_1 T_2 S_2 e_i = T_1 e_i$ for $i > m$ as well, since both sides are 0. Since $S_1 T_2 S_2$ and T_1 agree on the basis $e_1, \ldots, e_n$, we have $S_1 T_2 S_2 = T_1$ by the linear map lemma.

$\Leftarrow$ Suppose there exist unitary operators $S_1, S_2 \in \text{End}(V)$ such that $T_1 = S_1 T_2 S_2$. By 7.5(d) and 7.53(b), we have

$$T_1^* T_1 = (S_1 T_2 S_2)^* (S_1 T_2 S_2) = S_2^* T_2^* S_1^* S_1 T_2 S_2 = S_2^* T_2^* T_2 S_2.$$

Moreover, 7.53(c) shows that $S_2^* = S_2^{-1}$. Therefore, $T_1^* T_1 v = \lambda v$ if and only if $T_2^* T_2 (S_2 v) = \lambda (S_2 v)$ for all $\lambda \in \mathbb{F}$ and $v \in V$. Hence, S_2 maps $E(\lambda, T_1^* T_1)$ onto $E(\lambda, T_2^* T_2)$ for all $\lambda \in \mathbb{F}$. Since S_2 is invertible, $\dim E(\lambda, T_1^* T_1) = \dim E(\lambda, T_2^* T_2)$ for all $\lambda \in \mathbb{F}$. By definition, the singular values of T_1 are the square roots of the eigenvalues of $T_1^* T_1$, each eigenvalue λ repeated $\dim E(\lambda, T_1^* T_1)$ times, and similarly for T_2. Therefore, T_1 and T_2 have the same singular values. □

Solution 7E.14 If $v = 0$, the inequality $s_n \|v\| \leq \|Tv\|$ holds trivially. Now assume $v \neq 0$. Let $u = \frac{v}{\|v\|}$, where $\|u\| = 1$ by construction. By Exercise 7E4, $\|Tu\| \geq s_n$. Then

$$\|Tv\| = \|T(\|v\| u)\| = \|v\| \, \|Tu\| \geq \|v\| s_n = s_n \|v\|. \qquad \square$$

Solution 7E.15 Let λ be an eigenvalue of T with corresponding eigenvector $v \neq 0$, so $Tv = \lambda v$. Let $u = \frac{v}{\|v\|}$, so $\|u\| = 1$. Then $Tu = \frac{1}{\|v\|} Tv = \frac{1}{\|v\|} \lambda v = \lambda u$. Hence, $\|Tu\| = |\lambda|$. By Exercise 7E4, $s_n \leq \|Tu\| \leq s_1$. Therefore, $s_1 \geq |\lambda| \geq s_n$. □

Solution 7E.16 Let $s_1, \ldots, s_m$ be the positive singular values of T. By the singular value decomposition, there exist orthonormal lists $e_1, \ldots, e_m$ in V and $f_1, \ldots, f_m$ in W such that

$$Tv = \sum_{i=1}^{m} s_i \langle v, e_i \rangle f_i$$

for all $v \in V$. By 7.75,

$$T^* w = \sum_{i=1}^{m} s_i \langle w, f_i \rangle e_i, \quad T^\dagger w = \sum_{i=1}^{m} \frac{1}{s_i} \langle w, f_i \rangle e_i$$

for all $w \in W$. By Exercise 7E9, $s_1, \ldots, s_m$ are the positive singular values of T^* as well. Applying 7.75 to T^* with the orthonormal lists $f_1, \ldots, f_m$ in W and $e_1, \ldots, e_m$ in V, we get

$$(T^*)^\dagger v = \sum_{i=1}^{m} \frac{1}{s_i} \langle v, e_i \rangle f_i$$

for all $v \in V$. Since each $s_i \in \mathbb{R}$, for all $w \in W$ and $v \in V$,

$$\langle T^\dagger w, v\rangle = \sum_{i=1}^{m} \frac{1}{s_i}\langle w, f_i\rangle\langle e_i, v\rangle = \left\langle w, \sum_{i=1}^{m} \frac{1}{s_i}\langle v, e_i\rangle f_i \right\rangle.$$

By the definition of the adjoint, this shows that

$$(T^\dagger)^* v = \sum_{i=1}^{m} \frac{1}{s_i}\langle v, e_i\rangle f_i$$

for all $v \in V$. This matches the expression for $(T^*)^\dagger$ obtained earlier, so $(T^*)^\dagger = (T^\dagger)^*$. □

Solution 7E.17 ⇒ Suppose T is self-adjoint. Then $T = T^*$. By Exercise 7E16, $T^\dagger = (T^*)^\dagger = (T^\dagger)^*$, so $T^\dagger$ is self-adjoint.
⇐ Suppose $T^\dagger$ is self-adjoint. Applying ⇒ to $T^\dagger$ shows that $(T^\dagger)^\dagger$ is self-adjoint. By Exercise 6C23, $(T^\dagger)^\dagger = T$, so T is self-adjoint. □

§7F

Solution 7F.1 By 7.87(d), $\|S\| = \|(S - T) + T\| \le \|S - T\| + \|T\|$. Similarly, $\|T\| = \|(T - S) + S\| \le \|T - S\| + \|S\|$. By 7.87(c) with $\lambda = -1$, $\|T - S\| = \|-(S - T)\| = \|S - T\|$. Combining the inequalities yields the desired result. □

Solution 7F.2

Appendix. *An alternate proof is given at the end of this solution.*

If $V = \{0\}$, then T has no eigenvalues, so we adopt the convention that the maximum over the empty set is 0. In this case, the only operator on V is $T = 0$, so $\|T\| = 0$ by 7.87(b), and the equality holds. Now assume $V \ne \{0\}$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T, where each eigenvalue λ of T appears in the list as many times as $\dim E(\lambda, T)$. As in the solution to Exercise 7E7, $n = \dim V \ge 1$. By Exercise 7E7, the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$ sorted in decreasing order. Each eigenvalue λ of T appears in the list $\lambda_1, \ldots, \lambda_n$ at least once since $\dim E(\lambda, T) \ge 1$. We then have

$$\max_{1 \le k \le n} |\lambda_k| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

By 7.88(a), $\|T\|$ is the largest singular value of T, which is the left-hand side of the equality in the previous line.

Alternate proof. Civil Service Pigeon. If $V = \{0\}$, the statement holds trivially.

Otherwise, since T is self-adjoint or complex normal, the spectral theorem guarantees an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$. For any vector $v = \sum_{i=1}^{n} c_i e_i \in V$,

$$\|Tv\|^2 = \sum_{i=1}^{n} |\lambda_i|^2 |c_i|^2 \le \left(\max_i |\lambda_i|\right)^2 \sum_{i=1}^{n} |c_i|^2 = \left(\max_i |\lambda_i|\right)^2 \|v\|^2,$$

which implies $\|T\| \le \max_i |\lambda_i|$.

Conversely, letting j be an index where the maximum is attained, $\|Te_j\| = |\lambda_j|\|e_j\|$, so

$$\|T\| \ge \frac{\|Te_j\|}{\|e_j\|} = |\lambda_j| = \max_i |\lambda_i|.$$

Therefore,

$$\|T\| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

□

Solution 7F.3 $\Rightarrow$ Define $S = \|T\|^2 I - T^*T$. By 7.5, T^*T is self-adjoint. Since $\|T\|^2 \in \mathbb{R}$, $(\|T\|^2 I)^* = \|T\|^2 I$. Hence, S is self-adjoint. For every $u \in V$, 7.88(c) gives

$$\langle Su, u\rangle = \langle(\|T\|^2 I - T^*T)u, u\rangle = \|T\|^2\|u\|^2 - \|Tu\|^2 \ge 0.$$

Hence, S is a positive operator. By hypothesis, $\langle Sv, v\rangle = \langle(\|T\|^2 I - T^*T)v, v\rangle = \|T\|^2\|v\|^2 - \|Tv\|^2 = 0$. Since S is positive, $\langle Sv, v\rangle = 0$ implies $Sv = 0$ by 7.43. Therefore, $(\|T\|^2 I - T^*T)v = 0$, which gives $T^*Tv = \|T\|^2 v$.
$\Leftarrow$ Suppose $T^*Tv = \|T\|^2 v$. Then $\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = \|T\|^2\langle v, v\rangle = \|T\|^2\|v\|^2$. Taking square roots gives $\|Tv\| = \|T\|\,\|v\|$. □

Solution 7F.4 Suppose $u \in V$ with $\langle u, v\rangle = 0$. By Exercise 7F3, $T^*Tv = \|T\|^2 v$. By the definition of the adjoint, the previous line, and 6.6(e), since $\|T\|^2 \in \mathbb{R}$, we have

$$\langle Tu, Tv\rangle = \langle u, T^*Tv\rangle = \langle u, \|T\|^2 v\rangle = \|T\|^2\langle u, v\rangle = 0.$$ □

Solution 7F.5 By 7.87(a), $\|S\| \ge 0$. By 7.88(c) applied to S and T, for every $v \in V$, we have

$$\|STv\| \le \|S\|\,\|Tv\| \le \|S\|\,\|T\|\,\|v\|.$$

By 7.88(c), $\|ST\|$ is the smallest number c such that $\|STv\| \le c\|v\|$ for all $v \in V$. Hence, $\|ST\| \le \|S\|\,\|T\|$. □

Solution 7F.6 The statement is false. Suppose $\dim V \ge 2$. By 6.35, there exists an orthonormal basis $e_1, e_2, \ldots, e_{\dim V}$ of V. Define the maps S and T by

$$Sv = \langle v, e_1\rangle e_1, \quad Tv = \langle v, e_1\rangle e_2$$

for all $v \in V$. By 6.6(a), $S, T \in \mathrm{End}(V)$. Then $STv = \langle v, e_1\rangle Se_2 = \langle v, e_1\rangle\langle e_2, e_1\rangle e_1 = 0$ for all $v \in V$. On the other hand, $TSv = \langle v, e_1\rangle Te_1 = \langle v, e_1\rangle e_2 = Tv$ for all $v \in V$. Therefore, $ST = 0$ and $TS = T$. Moreover, $T \ne 0$ because $Te_1 = e_2 \ne 0$. By 7.87(a) and (b), we have $\|ST\| = 0 < \|TS\| = \|T\|$. □

Solution 7F.7 We verify the metric axioms for d:

- **Nonnegativity and identity**: Nonnegativity follows from 7.87(a). By 7.87(b), $d(S,T) = \|S-T\| = 0$ if and only if $S - T = 0$, i.e., $S = T$.
- **Symmetry**: Since $T-S = (-1)(S-T)$, 7.87(c) implies $\|T-S\| = \|(-1)(S-T)\| = \|S-T\|$, so $d(S,T) = d(T,S)$.
- **Triangle inequality**: For any $R \in \operatorname{Hom}(V,W)$, $S - R = (S-T) + (T-R)$, so by 7.87(d),

$$d(S,R) = \|S-R\| \le \|S-T\| + \|T-R\| = d(S,T) + d(T,R).$$

Therefore, $(\operatorname{Hom}(V,W), d)$ is a metric space. □

Solution 7F.8

(a) By 3.65, it suffices to show that T is injective. Let $v \in V$ be such that $Tv = 0$. Then $(I-T)v = v$. By 7.88(c), $\|I-T\|\,\|v\| \ge \|(I-T)v\| = \|v\|$. If $v \neq 0$, then dividing both sides by $\|v\|$ gives $\|I-T\| \ge 1$, which contradicts the assumption that $\|I-T\| < 1$. Therefore, $v = 0$, and T is injective.

(b) Note that the hypothesis $\|S-T\| < 1/\|S^{-1}\|$ requires that $\|S^{-1}\| \neq 0$, so $\|S^{-1}\| > 0$ by 7.87(a). By 3.8, $I - S^{-1}T = S^{-1}(S-T)$. By Exercise 7F5,

$$\|I - S^{-1}T\| \le \|S^{-1}\|\,\|S-T\|.$$

Since $\|S-T\| < 1/\|S^{-1}\|$ and $\|S^{-1}\| > 0$, $\|I - S^{-1}T\| < 1$. By part (a), $S^{-1}T$ is invertible, so $T = S(S^{-1}T)$ is invertible.

□

Solution 7F.9

Manual Remark. *Note that this exercise assumes $V \neq \{0\}$, for otherwise, $\|T-S\| = 0$ for every $S \in \operatorname{End}(V)$.*

Since $(0,\varepsilon)$ contains infinitely many real numbers, and T has at most $\dim V$ distinct eigenvalues by 5.12, there exists a real number $\lambda \in (0,\varepsilon)$ that is not an eigenvalue of T. Consider the operator $S = T - \lambda I$. By 5.7, S is invertible because λ is not an eigenvalue of T. Moreover, since I is an isometry, its singular values are all 1 by 7.69, so $\|I\| = 1$ by 7.88(a). Therefore, by 7.87(c), $\|T-S\| = \|\lambda I\| = |\lambda|\|I\| = \lambda$, where $|\lambda| = \lambda$ because $\lambda > 0$. Hence, $0 < \|T-S\| = \lambda < \varepsilon$. □

Solution 7F.10 By 3.65, T is not injective since it is not invertible. Then there exists a nonzero $u \in \ker T$. Let $U = (\operatorname{span}(u))^{\perp}$. Since $\dim \operatorname{span}(u) = 1$ and $\dim V > 1$, we have $\dim U = \dim V - 1 \ge 1$ by 6.51. Consider the operator $S \in \operatorname{End}(V)$ defined by $S = T + \frac{\varepsilon}{2} P_U$. Note that for every $w \in U$, $\langle w, u\rangle = 0$, so $u \in U^{\perp}$ and hence $P_U u = 0$ by 6.57(c). Then $Su = 0$, so S is not injective, and hence not invertible by 3.65. By 6.57(h), $\|P_U v\| \le \|v\|$ for all $v \in V$, which yields $\|P_U\| \le 1$ by 7.88(c). Since $\dim U \ge 1$, there exists a unit vector $x \in U$. By 6.57(b), $P_U x = x$, so $\|P_U\| \ge \|P_U x\| = \|x\| = 1$ by 7.88(b). Thus, $\|P_U\| = 1$. Finally, by 7.87(c),

$$\|T-S\| = \left\| -\frac{\varepsilon}{2} P_U \right\| = \frac{\varepsilon}{2}\|P_U\| = \frac{\varepsilon}{2},$$

so $0 < \|T - S\| < \varepsilon$. □

Solution 7F.11

Manual Remark. *Note that this exercise assumes $V \neq \{0\}$, for otherwise, $\|T - S\| = 0$ for every $S \in \mathrm{End}(V)$.*

Let $\dim V = n$. By Schur's theorem, since $\mathbb{F} = \mathbb{C}$, there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $\mathrm{Mat}(T)$ is upper triangular with respect to this basis with diagonal entries $\lambda_1, \ldots, \lambda_n$, which are the eigenvalues of T by 5.41, possibly with repetitions. Define $\delta_1 = \frac{\varepsilon}{2}$. For $k = 2, \ldots, n$, choose $\delta_k \in [0, \frac{\varepsilon}{2})$ such that

$$\lambda_k + \delta_k \notin \{\lambda_j + \delta_j : 1 \leq j < k\}.$$

Note that this construction is possible since $[0, \frac{\varepsilon}{2})$ contains infinitely many real numbers, and only $k - 1$ values are excluded for each $k \geq 2$. By the linear map lemma, we may construct an operator $D \in \mathrm{End}(V)$ by $De_k = \delta_k e_k$. Then $\mathrm{Mat}(D, (e_1, \ldots, e_n))$ is diagonal with diagonal entries $\delta_1, \ldots, \delta_n$, which are its eigenvalues by 5.41.

Let $S = T + D$. By 3.35, $\mathrm{Mat}(S, (e_1, \ldots, e_n))$ is upper triangular with diagonal entries $\lambda_1 + \delta_1, \ldots, \lambda_n + \delta_n$, which are distinct by construction. Since these are n distinct eigenvalues by 5.41, S is diagonalizable by 5.58. By the complex spectral theorem, D is normal since it has a diagonal matrix with respect to the orthonormal basis $e_1, \ldots, e_n$. By Exercise 7F2, and since $\delta_k \geq 0$ for all k,

$$\|T - S\| = \|D\| = \max_{1 \leq k \leq n} |\delta_k| = \max_{1 \leq k \leq n} \delta_k = \frac{\varepsilon}{2} < \varepsilon.$$

Then $0 < \|T - S\| = \frac{\varepsilon}{2} < \varepsilon$. □

Solution 7F.12 By 7.39, $\sqrt{T}$ is the unique positive square root of T, so T and $\sqrt{T}$ are self-adjoint by 7.38(b). By Exercise 7F2 and the nonnegativity of the eigenvalues of T and $\sqrt{T}$ by 7.38(b), we have

$$\|T\| = \max_{\lambda} \lambda, \quad \left\|\sqrt{T}\right\| = \max_{\mu} \mu,$$

where λ and μ range over all eigenvalues of T and $\sqrt{T}$, respectively.
By the proof of 7.39, for every eigenvalue λ of T, $\sqrt{\lambda}$ is an eigenvalue of $\sqrt{T}$. Moreover, if $\sqrt{T}v = \mu v$ for some nonzero $v \in V$, then $Tv = \mu^2 v$. Since $\mu \geq 0$, we have $\mu = \sqrt{\mu^2}$, and μ^2 is an eigenvalue of T since $v \neq 0$. Therefore, the eigenvalues of $\sqrt{T}$ are precisely the square roots of the eigenvalues of T, so

$$\left\|\sqrt{T}\right\| = \max_{\mu} \mu = \max_{\lambda} \sqrt{\lambda} = \sqrt{\max_{\lambda} \lambda} = \sqrt{\|T\|},$$

where the third equality follows from the fact that $\lambda \mapsto \sqrt{\lambda}$ is an increasing function on $[0, \infty)$. □

Solution 7F.13 If $V = \{0\}$, then $S = T = 0$, and the inequality holds trivially. Now assume $V \neq \{0\}$. Since S and T are positive, the Cauchy–Schwarz inequality and 7.88(c) give

$$0 \leq \langle Sv, v \rangle \leq \|S\|, \quad 0 \leq \langle Tv, v \rangle \leq \|T\|$$

for every unit vector $v \in V$. By 7.38(b), S and T are self-adjoint, so $S - T$ is also self-adjoint by 7.5(a) and (b). By Exercise 7F2, there exists an eigenvalue λ of $S - T$ such that $\|S - T\| = |\lambda|$. Let $u \in V$ be a unit eigenvector corresponding to λ, so $(S - T)u = \lambda u$. Then

$$\lambda = \lambda\|u\|^2 = \langle (S - T)u, u\rangle = \langle Su, u\rangle - \langle Tu, u\rangle \in [-\|T\|, \|S\|].$$

Thus, $|\lambda| = \|S - T\| \leq \max\{\|S\|, \|T\|\}$.
By Exercise 7F2 again, there exists an eigenvalue μ of S such that $\|S\| = |\mu|$. By 7.38(b), $\mu \geq 0$, so $\|S\| = |\mu| = \mu$. Let $w \in V$ be a unit eigenvector corresponding to μ, so $Sw = \mu w$. Then $\langle Sw, w\rangle = \mu\|w\|^2 = \mu = \|S\|$. Consequently,

$$\|S\| \leq \|S\| + \langle Tw, w\rangle = \langle (S + T)w, w\rangle \leq \|S + T\|,$$

where the last inequality follows from the Cauchy–Schwarz inequality and 7.88(c). By symmetry, we also have $\|T\| \leq \|S + T\|$. Therefore, $\max\{\|S\|, \|T\|\} \leq \|S + T\|$, and the desired inequality is proved. □

Solution 7F.14 Suppose $\dim U > \dim W$. By 6.51, 2.43, and 2.37,

$$\dim(U \cap W^\perp) = \dim U + \dim W^\perp - \dim(U + W^\perp) \geq \dim U - \dim W > 0.$$

Let $u \in U \cap W^\perp$ be a unit vector. Then $(P_U - P_W)u = P_U u - P_W u = u - 0 = u$ by 6.57(b) and (c). By 7.88(c), $\|P_U - P_W\| \geq \|(P_U - P_W)u\| = \|u\| = 1$, contradicting the assumption that $\|P_U - P_W\| < 1$. Therefore, our assumption that $\dim U > \dim W$ must be false. By symmetry, $\dim W > \dim U$ is also impossible. Hence, $\dim U = \dim W$. □

Solution 7F.15 For any $(z_1, z_2, z_3), (w_1, w_2, w_3) \in \mathbb{F}^3$,

$$\begin{aligned}\langle T(z_1, z_2, z_3), (w_1, w_2, w_3)\rangle &= \langle (z_3, 2z_1, 3z_2), (w_1, w_2, w_3)\rangle \\ &= z_3\overline{w_1} + 2z_1\overline{w_2} + 3z_2\overline{w_3} \\ &= \langle (z_1, z_2, z_3), (2w_2, 3w_3, w_1)\rangle.\end{aligned}$$

Therefore,

$$T^*(w_1, w_2, w_3) = (2w_2, 3w_3, w_1), \quad T^*T(z_1, z_2, z_3) = (4z_1, 9z_2, z_3).$$

Let $R \in \text{End}(\mathbb{F}^3)$ be defined by $R(z_1, z_2, z_3) = (2z_1, 3z_2, z_3)$. Then $\text{Mat}(R)$ with respect to the standard orthonormal basis e_1, e_2, e_3 of $\mathbb{F}^3$ is diagonal with nonnegative diagonal entries $2, 3, 1$. By 7.38(c), R is positive. Additionally, $R^2(z_1, z_2, z_3) = (4z_1, 9z_2, z_3) = T^*T(z_1, z_2, z_3)$, so $R^2 = T^*T$ and hence $R = \sqrt{T^*T}$ by 7.39. Moreover, R^{-1} exists and is given by

$$R^{-1}(z_1, z_2, z_3) = \left(\frac{z_1}{2}, \frac{z_2}{3}, z_3\right).$$

Define $S \in \text{End}(\mathbb{F}^3)$ by $S = TR^{-1}$. Then

$$S(z_1, z_2, z_3) = T(R^{-1}(z_1, z_2, z_3)) = T\left(\frac{z_1}{2}, \frac{z_2}{3}, z_3\right) = (z_3, z_1, z_2).$$

Observe that $Se_1 = e_2$, $Se_2 = e_3$, and $Se_3 = e_1$. Since e_2, e_3, e_1 is an orthonormal basis of $\mathbb{F}^3$, S is unitary by 7.53(d). Because $S = TR^{-1}$, we have $T = SR = S\sqrt{T^*T}$. □

Solution 7F.16 If $V = \{0\}$, every operator on V is positive, so any $\delta > 0$ works. Now assume $V \neq \{0\}$. By 7.38(c), there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $Se_k = \lambda_k e_k$ with $\lambda_k \geq 0$ for all $k = 1, \ldots, n$. Then each λ_k is an eigenvalue of S because $e_k \neq 0$, and since S is invertible, $\lambda_k > 0$ for all $k = 1, \ldots, n$ by 5.7. Let

$$\delta = \min_{1 \leq k \leq n} \lambda_k > 0.$$

Observe that for any $v \in V$ and $k = 1, \ldots, n$, we have

$$\langle Sv, e_k \rangle = \langle v, Se_k \rangle = \overline{\lambda_k} \langle v, e_k \rangle = \lambda_k \langle v, e_k \rangle,$$

since S is self-adjoint and $\lambda_k \in \mathbb{R}$ for all k. By 6.30(b) and (c),

$$\langle Sv, v \rangle = \sum_{k=1}^{n} \langle Sv, e_k \rangle \overline{\langle v, e_k \rangle} = \sum_{k=1}^{n} \lambda_k |\langle v, e_k \rangle|^2 \geq \delta \sum_{k=1}^{n} |\langle v, e_k \rangle|^2 = \delta \|v\|^2.$$

Let $T \in \text{End}(V)$ be self-adjoint with $\|S - T\| < \delta$. By the Cauchy–Schwarz inequality and 7.88(c),

$$\langle Tv, v \rangle = \langle Sv, v \rangle - \langle (S - T)v, v \rangle \geq \langle Sv, v \rangle - \|S - T\| \, \|v\|^2 \geq (\delta - \|S - T\|)\|v\|^2 \geq 0.$$

Since T is self-adjoint and $\langle Tv, v \rangle \geq 0$ for all $v \in V$, it follows that T is positive. □

Solution 7F.17 Observe that $\|\alpha\| = \sqrt{\langle \alpha, \alpha \rangle} = \sqrt{\alpha \overline{\alpha}} = |\alpha|$ for all $\alpha \in \mathbb{F}$. For any $v \in V$, the Cauchy–Schwarz inequality gives

$$|\phi_u(v)| = |\langle v, u \rangle| \leq \|v\| \, \|u\|.$$

By 7.88(c), $\|\phi_u\| \leq \|u\|$.
If $u = 0$, then $0 \leq \|\phi_u\| \leq \|u\| = 0$ by 7.87(a). For $u \neq 0$, consider $w = \frac{u}{\|u\|}$. Since $\|w\| = 1$, 7.88(b) gives $\|\phi_u\| \geq |\phi_u(w)|$, so

$$\|\phi_u\| \geq |\phi_u(w)| = |\langle w, u \rangle| = \left| \left\langle \frac{u}{\|u\|}, u \right\rangle \right| = \frac{\|u\|^2}{\|u\|} = \|u\|.$$

Hence, $\|\phi_u\| = \|u\|$. □

Solution 7F.18

Axler Note. *Here $e_1, \ldots, e_n$ is an arbitrary orthonormal basis of V, not necessarily connected with a singular value decomposition of T. If $s_1, \ldots, s_n$ is the list of singular values of T, then the right side of the inequality above equals $(s_1^2 + \cdots + s_n^2)^{1/2}$, as was shown in Exercise 7E11(a).*

Let $s_1 \geq \cdots \geq s_n$ be the singular values of T.

(a) Observe that $\max\{\|Te_1\|, \dots, \|Te_n\|\} \le \|T\|$ by 7.88(b) applied to each e_k. By Exercise 7E11(a) and 7.88(a),

$$\sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{n} s_k^2 \ge s_1^2 = \|T\|^2.$$

Taking square roots gives the desired inequality.

(b) Note that $\|T\| = (\sum_{k=1}^{n} \|Te_k\|^2)^{1/2}$ if and only if $\|T\|^2 = \sum_{k=1}^{n} \|Te_k\|^2$, since both sides are nonnegative.

$$\begin{aligned}
\|T\|^2 = \sum_{k=1}^{n} \|Te_k\|^2 &\iff s_1^2 = \sum_{k=1}^{n} s_k^2 \\
&\iff \sum_{k=2}^{n} s_k^2 = 0 \\
&\iff s_2 = \cdots = s_n = 0 \\
&\iff T \text{ has at most one positive singular value} \\
&\iff \dim \operatorname{im} T \le 1,
\end{aligned}$$

where the first equivalence follows from 7.88(a) and Exercise 7E11(a); the third equivalence from $s_k^2 \ge 0$ for all k; the fourth equivalence from the decreasing order of the singular values; and the fifth equivalence from 7.68(b). □

Solution 7F.19

Axler Note. *This formula for $\|T^*T\|$ leads to the important subject of C^*-algebras.*

Manual Remark. *Alternatively, if $T \neq 0$ and*

$$Tv = \sum_{k=1}^{m} s_k \langle v, e_k \rangle f_k$$

is a singular value decomposition of T by the singular value decomposition, then 7.75 gives

$$T^*Tv = \sum_{k=1}^{m} s_k^2 \langle v, e_k \rangle e_k,$$

so Exercise 7E8(c) and 7.88(a) give $\|T^*T\| = s_1^2 = \|T\|^2$.

$\le$ By 7.88(b), there exists $v \in V$ with $\|v\| = 1$ such that $\|Tv\| = \|T\|$. By the definition of the adjoint, the Cauchy–Schwarz inequality, and 7.88(c), we have

$$\|T\|^2 = \|Tv\|^2 = \langle Tv, Tv \rangle = \langle v, T^*Tv \rangle \le \|v\| \, \|T^*Tv\| \le \|T^*T\| \, \|v\|^2 = \|T^*T\|.$$

$\ge$ By Exercise 7F5 and 7.91, $\|T^*T\| \le \|T^*\| \, \|T\| = \|T\|^2$. □

Solution 7F.20 Let $n = \dim V$ and let $S = T^*T$. We first show that $(T^k)^*T^k = S^k$ for all positive integers k. The base case $k = 1$ is trivial since $(T^1)^*T^1 = T^*T = S$. Now suppose $(T^{k-1})^*T^{k-1} = S^{k-1}$ for $k \ge 2$. Since T is normal, S commutes with T, and

hence with T^{k-1}. By 7.5(d) applied to T and T^{k-1} and the induction hypothesis, we have

$$(T^k)^*T^k = (TT^{k-1})^*T^k = (T^{k-1})^*(T^*T)T^{k-1} = (T^{k-1})^*ST^{k-1} = (T^{k-1})^*T^{k-1}S = S^{k-1}S = S^k,$$

so $(T^k)^*T^k = S^k$ for all positive integers k.
Observe that S is positive by 7.64(a), so by 7.38(c), $\mathrm{Mat}(S)$ is diagonal with entries $\lambda_1, \ldots, \lambda_n \geq 0$ with respect to some orthonormal basis of V. By 3.43, $\mathrm{Mat}(S^k)$ is also diagonal with entries $\lambda_1^k, \ldots, \lambda_n^k$ with respect to the same orthonormal basis of V, which are exactly the eigenvalues of S^k by 5.41, so S^k is positive by 7.38(c). By Exercise 7F2 applied to S^k and S, the nonnegativity of the eigenvalues of S^k, and the fact that $x \mapsto x^k$ is increasing on $[0, \infty)$, we have

$$\|S^k\| = \max_{1 \leq j \leq n} \lambda_j^k = \left(\max_{1 \leq j \leq n} \lambda_j \right)^k = \|S\|^k.$$

The above computation and Exercise 7F19 applied to T^k and T then give $\|T^k\|^2 = \|S\|^k = \|T\|^{2k}$, and taking the nonnegative square roots of both sides yields $\|T^k\| = \|T\|^k$. □

Solution 7F.21 Suppose the norm on $\mathrm{Hom}(V, W)$ comes from an inner product. By 6.35, there exist orthonormal bases $e_1, \ldots, e_{\dim V}$ of V and $f_1, \ldots, f_{\dim W}$ of W. Since $\dim V > 1$ and $\dim W > 1$, we may take the orthonormal lists e_1, e_2 in V and f_1, f_2 in W. Consider the maps S and R defined by

$$Sv = \langle v, e_1 \rangle f_1, \quad Rv = \langle v, e_2 \rangle f_2.$$

Note that $S, R \in \mathrm{Hom}(V, W)$ by 6.6(a). By the Cauchy–Schwarz inequality,

$$\|Sv\| = |\langle v, e_1 \rangle| \, \|f_1\| \leq \|v\| \leq 1$$

for $\|v\| \leq 1$. Since $\|Se_1\| = \|f_1\| = 1$, we have $\|S\| = 1$. Similarly, $\|R\| = 1$, using e_2 and f_2. By 6.24 and Bessel's inequality, we have for any $v \in V$ that

$$\|(S \pm R)v\|^2 = |\langle v, e_1 \rangle|^2 + |\langle v, e_2 \rangle|^2 \leq \|v\|^2.$$

It follows that $\|(S \pm R)v\| \leq \|v\| \leq 1$ for all $v \in V$ with $\|v\| \leq 1$, so $\|S \pm R\| \leq 1$. Because $(S \pm R)e_1 = f_1$, it follows that $\|S + R\| \geq 1$ and $\|S - R\| \geq 1$. Hence, $\|S + R\| = \|S - R\| = 1$. Consequently,

$$\|S + R\|^2 + \|S - R\|^2 = 1^2 + 1^2 = 2$$

but

$$2\|S\|^2 + 2\|R\|^2 = 2 \cdot 1^2 + 2 \cdot 1^2 = 4.$$

The parallelogram equality requires that $\|S + R\|^2 + \|S - R\|^2 = 2\|S\|^2 + 2\|R\|^2$, which is not the case here. Therefore, the norm on $\mathrm{Hom}(V, W)$ does not come from an inner product. □

Solution 7F.22 Let m be the number of positive singular values of T, so $s_{m+1} = \cdots = s_n = 0$. By the singular value decomposition, there exist orthonormal lists $e_1, \ldots, e_m$ in V and $f_1, \ldots, f_m$ in W such that

$$Tv = \sum_{j=1}^{m} s_j \langle v, e_j \rangle f_j$$

for all $v \in V$. Extend $e_1, \ldots, e_m$ to an orthonormal basis $e_1, \ldots, e_n$ of V by 6.35 if $m = 0$, or by 6.36 if $m > 0$. By 6.24,

$$\|Tv\|^2 = \sum_{j=1}^{m} s_j^2 |\langle v, e_j \rangle|^2 + \sum_{j=m+1}^{n} 0 \cdot |\langle v, e_j \rangle|^2 = \sum_{j=1}^{n} s_j^2 |\langle v, e_j \rangle|^2.$$

$\leq$ Consider the subspace $X = \operatorname{span}(e_{n-k+1}, \ldots, e_n)$ of V. By 6.25, $\dim X = k$. For every $x \in X$ and $j \leq n-k$, we have $\langle x, e_j \rangle = 0$ by the orthonormality of $e_1, \ldots, e_n$. By the above calculation,

$$\|Tx\|^2 = \sum_{j=n-k+1}^{n} s_j^2 |\langle x, e_j \rangle|^2 \leq s_{n-k+1}^2 \sum_{j=n-k+1}^{n} |\langle x, e_j \rangle|^2 = s_{n-k+1}^2 \|x\|^2,$$

where the last equality follows from 6.30(b). By 7.88(c), $\|T|_X\| \leq s_{n-k+1}$.

$\geq$ Let U be any subspace of V with $\dim U = k$. Let $Y = \operatorname{span}(e_1, \ldots, e_{n-k+1})$, so $\dim Y = n-k+1$ by 6.25. By 2.37, $\dim(U+Y) \leq n$. By 2.43,

$$\dim(U \cap Y) = \dim U + \dim Y - \dim(U+Y) \geq k + (n-k+1) - n = 1.$$

Let $u \in U \cap Y$ be nonzero. For all $j > n-k+1$, we have $\langle u, e_j \rangle = 0$ by the orthonormality of $e_1, \ldots, e_n$. By the above calculation,

$$\|Tu\|^2 = \sum_{j=1}^{n-k+1} s_j^2 |\langle u, e_j \rangle|^2 \geq s_{n-k+1}^2 \sum_{j=1}^{n-k+1} |\langle u, e_j \rangle|^2 = s_{n-k+1}^2 \|u\|^2,$$

where the last equality follows from 6.30(b). Taking square roots yields $\|Tu\| \geq s_{n-k+1}\|u\|$. By 7.88(c), $\|T|_U\| \, \|u\| \geq \|Tu\| \geq s_{n-k+1}\|u\|$, so $\|T|_U\| \geq s_{n-k+1}$ by dividing both sides by $\|u\| > 0$ since $u \neq 0$. Since every U with $\dim U = k$ satisfies $\|T|_U\| \geq s_{n-k+1}$, and we have found an X with $\dim X = k$ such that $\|T|_X\| \leq s_{n-k+1}$, the minimum exists and is attained at X, and therefore equals s_{n-k+1}. □

Solution 7F.23

Manual Remark. *For reference, the metrics on V and W that arise from the norms are $d_V(v_1, v_2) = \|v_1 - v_2\|$ and $d_W(w_1, w_2) = \|w_1 - w_2\|$. With these metrics, uniform continuity of T means that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $v_1, v_2 \in V$, if $\|v_1 - v_2\| < \delta$, then $\|Tv_1 - Tv_2\| < \varepsilon$.*

Let $\varepsilon > 0$ be given. Define

$$\delta = \frac{\varepsilon}{\|T\| + 1}.$$

Note that $\delta > 0$ since $\|T\| \geq 0$ by 7.87(a). Suppose $v_1, v_2 \in V$ such that $\|v_1 - v_2\| < \delta$. By 7.88(c),

$$\|Tv_1 - Tv_2\| = \|T(v_1 - v_2)\| \leq \|T\| \, \|v_1 - v_2\| \leq \|T\|\delta = \|T\| \frac{\varepsilon}{\|T\| + 1} < \varepsilon.$$

Hence, T is uniformly continuous. □

Solution 7F.24 Since T is invertible, $T \neq 0$ so $\|T\| \neq 0$ by 7.87(b). Let $S = \frac{T}{\|T\|}$, so $S^{-1} = \|T\|T^{-1}$. By 7.87(c), $\|S\| = 1$.
$\Rightarrow$ Suppose $\|T^{-1}\| = \|T\|^{-1}$. By 7.87(c), $\|S^{-1}\| = \|T^{-1}\|\|T\| = 1$. Then for every $v \in V$, we have by 7.88(c) that

$$\|v\| = \|S^{-1}Sv\| \leq \|S^{-1}\|\,\|Sv\| = \|Sv\| \leq \|S\|\,\|v\| = \|v\|,$$

so $\|Sv\| = \|v\|$. Therefore, S is an invertible isometry, so S is unitary.

$\Leftarrow$ Suppose S is unitary. Then $S^{-1} = S^*$ by 7.53(c), and hence is unitary by 7.53(f). Thus, $\|S^{-1}\| = 1$ by 7.88(b), so $\|T\|\,\|T^{-1}\| = \|S^{-1}\| = 1$ by 7.87(c). Hence, $\|T^{-1}\| = \|T\|^{-1}$. □

Solution 7F.25 Observe that $\langle Tv, w\rangle = \langle\langle v, u\rangle x, w\rangle = \langle v, u\rangle\langle x, w\rangle$ for all $v, w \in V$. Then $\langle v, u\rangle\langle x, w\rangle = \langle v, \overline{\langle x, w\rangle}u\rangle = \langle v, \langle w, x\rangle u\rangle$ for all $v, w \in V$, so $T^*w = \langle w, x\rangle u$ for all $w \in V$. Therefore,

$$T^*Tv = T^*(\langle v, u\rangle x) = \langle v, u\rangle T^*x = \langle v, u\rangle\langle x, x\rangle u = \|x\|^2\langle v, u\rangle u$$

for all $v \in V$. Define $R \in \text{End}(V)$ by $Rv = \frac{\|x\|}{\|u\|}\langle v, u\rangle u$ for all $v \in V$. Since $\frac{\|x\|}{\|u\|} \geq 0$, $\langle Rv, w\rangle = \frac{\|x\|}{\|u\|}\langle v, u\rangle\langle u, w\rangle = \langle v, Rw\rangle$ for all $v, w \in V$, so R is self-adjoint. Moreover,

$$\langle Rv, v\rangle = \frac{\|x\|}{\|u\|}\langle v, u\rangle\langle u, v\rangle = \frac{\|x\|}{\|u\|}|\langle v, u\rangle|^2 \geq 0,$$

so R is positive. Finally,

$$R^2v = R\left(\frac{\|x\|}{\|u\|}\langle v, u\rangle u\right) = \frac{\|x\|}{\|u\|}\left\langle\frac{\|x\|}{\|u\|}\langle v, u\rangle u, u\right\rangle u = \frac{\|x\|^2}{\|u\|^2}\langle v, u\rangle\langle u, u\rangle u = \|x\|^2\langle v, u\rangle u = T^*Tv$$

for all $v \in V$. Then R is a positive square root of T^*T, so $R = \sqrt{T^*T}$ by 7.39. Hence,

$$\sqrt{T^*T}\,v = Rv = \frac{\|x\|}{\|u\|}\langle v, u\rangle u$$

for all $v \in V$. □

Solution 7F.26 Since $\sqrt{T^*T}$ is positive, it is self-adjoint. Then for any $v \in V$, we have

$$\left\|\sqrt{T^*T}\,v\right\|^2 = \langle T^*Tv, v\rangle = \|Tv\|^2.$$

Hence, $\ker\sqrt{T^*T} = \ker T$. By the polar decomposition of T, the existence of such S is guaranteed. As such, it remains to show that S is unique if and only if T is invertible.
$\Rightarrow$ Suppose T is invertible. By the above, $\ker\sqrt{T^*T} = \ker T = \{0\}$, so $\sqrt{T^*T}$ is injective, and hence invertible by 3.65. Moreover, if $S_1, S_2 \in \text{End}(V)$ are unitary operators such that $T = S_1\sqrt{T^*T} = S_2\sqrt{T^*T}$, then right-multiplying by $(\sqrt{T^*T})^{-1}$ gives $S_1 = S_2$, so S is unique.

$\Leftarrow$ Suppose T is not invertible. By 3.65, $\ker T \neq \{0\}$. By the polar decomposition of T, there exists a unitary operator S such that $T = S\sqrt{T^*T}$. By 6.35, there exists an orthonormal basis $e_1, \ldots, e_m$ of $\ker T$, where $m \geq 1$. By 6.36, we may extend this to an orthonormal basis $e_1, \ldots, e_n$ of V. By the linear map lemma, there exists $R \in \operatorname{End}(V)$ such that $Re_j = -e_j$ for $j \leq m$ and $Re_j = e_j$ for $j > m$. Because $Re_1, \ldots, Re_n$ is an orthonormal basis of V, R is unitary by 7.53(d). Moreover, self-adjointness of $\sqrt{T^*T}$ gives

$$\operatorname{im}\sqrt{T^*T} = (\ker\sqrt{T^*T})^\perp = (\ker T)^\perp$$

by 7.6(b). Now, if $u \in (\ker T)^\perp$, then $\langle u, e_j\rangle = 0$ for $j \leq m$. By 6.30(a), we may write

$$u = \sum_{j=m+1}^{n} \langle u, e_j\rangle e_j,$$

so $Ru = u$. Consequently, $R\sqrt{T^*T} = \sqrt{T^*T}$, so $SR\sqrt{T^*T} = S\sqrt{T^*T} = T$. By 7.5(d) and 7.53(b),

$$(SR)^*(SR) = R^*S^*SR = I, \quad (SR)(SR)^* = SRR^*S^* = I.$$

Hence, SR is unitary. Lastly, injectivity of S shows that $Se_1 \neq 0$, so $SRe_1 = -Se_1 \neq Se_1$. Therefore, S and SR are distinct unitary operators such that $T = S\sqrt{T^*T} = SR\sqrt{T^*T}$, and S is not unique. □

Solution 7F.27

(a) Since $Se_j = f_j$ for all j, S is unitary by 7.53(d). By assumption,

$$(T - S)v = \sum_{j=1}^{n}(s_j - 1)\langle v, e_j\rangle f_j$$

for any $v \in V$. Define

$$M = \max_{1\leq j\leq n} |s_j - 1|.$$

By 6.24 and 6.30(b),

$$\|(T - S)v\|^2 = \sum_{j=1}^{n}|s_j - 1|^2|\langle v, e_j\rangle|^2 \leq M^2\sum_{j=1}^{n}|\langle v, e_j\rangle|^2 = M^2\|v\|^2$$

for all $v \in V$. Taking square roots and applying 7.88(c) gives $\|T - S\| \leq M$. Choose k such that $|s_k - 1| = M$. Since $\|(T - S)e_k\| = |s_k - 1| = M$ and $\|e_k\| = 1$, 7.88(b) gives $\|T - S\| \geq M$. Therefore, $\|T - S\| = M$.

(b) Suppose $E \in \operatorname{End}(V)$ is unitary. Then it is an isometry, so $\|Ee_j\| = \|e_j\| = 1$ for all j. Moreover, since $Te_j = s_j f_j$ and $s_j \geq 0$ for all j, we have $\|Te_j\| = s_j$. Because $\|e_j\| = 1$, 7.88(b) and the reverse triangle inequality give

$$\|T - E\| \geq \|(T - E)e_j\| \geq \big|\|Te_j\| - \|Ee_j\|\big| = |s_j - 1|.$$

Since this holds for all j, taking the maximum over j gives

$$\|T - E\| \geq \max_{1\leq j\leq n} |s_j - 1| = \|T - S\|.$$ □

Solution 7F.28 By the polar decomposition of T^*, there exists a unitary operator $U \in \text{End}(V)$ such that $T^* = U\sqrt{(T^*)^*T^*}$. By 7.5(c), $(T^*)^* = T$, so $T^* = U\sqrt{TT^*}$. Since $\sqrt{TT^*}$ is positive, it is self-adjoint, so $\sqrt{TT^*} = (\sqrt{TT^*})^*$. By 7.5(c), (d), and taking adjoints, $T = (U\sqrt{TT^*})^* = \sqrt{TT^*}\,U^*$. By 7.53(f), U^* is unitary since U is unitary. Setting $S = U^*$ gives the desired result. □

Solution 7F.29

(a) By the polar decomposition of T, there exists a unitary operator $S \in \text{End}(V)$ such that $T = S\sqrt{T^*T}$. By 7.5(d), $T^* = (\sqrt{T^*T})^*S^*$. Since $\sqrt{T^*T}$ is positive, it is self-adjoint, so $(\sqrt{T^*T})^* = \sqrt{T^*T}$, and hence $T^* = \sqrt{T^*T}\,S^*$. By the definition of the square root,

$$TT^* = S\sqrt{T^*T}\left(\sqrt{T^*T}\,S^*\right) = ST^*TS^*.$$

(b) Since T^*T is positive, 7.38(c) gives an orthonormal basis $e_1, \ldots, e_n$ of V and nonnegative numbers $\mu_1, \ldots, \mu_n$ such that $T^*Te_k = \mu_k e_k$ for each k. By 7.53(b), $S^*S = I$ since S is unitary. Therefore,

$$TT^*(Se_k) = ST^*TS^*(Se_k) = ST^*Te_k = S(\mu_k e_k) = \mu_k(Se_k).$$

Since S is invertible, $Se_1, \ldots, Se_n$ is also a basis of V. By Lemma 1 of Section 7E applied to T with the basis $e_1, \ldots, e_n$, the singular values of T are $\sqrt{\mu_1}, \ldots, \sqrt{\mu_n}$ sorted into decreasing order. By 7.5(c), $(T^*)^*T^* = TT^*$, so $(T^*)^*T^*(Se_k) = TT^*(Se_k) = \mu_k(Se_k)$ for each k. By Lemma 1 of Section 7E applied to T^* with the basis $Se_1, \ldots, Se_n$, the singular values of T^* are $\sqrt{\mu_1}, \ldots, \sqrt{\mu_n}$ sorted into decreasing order, which are the same as those of T.

□

Solution 7F.30 Since R is positive, it is self-adjoint. By 7.5(d) and 7.53(b), $T^*T = RS^*SR = R^2$. Then R is a positive square root of T^*T, so $R = \sqrt{T^*T}$ by 7.39. □

Solution 7F.31 By the complex spectral theorem, there exists an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T, and scalars $\lambda_1, \ldots, \lambda_n$ such that $Te_k = \lambda_k e_k$ for each k. By 7.21(e),

$$T^*Te_k = \lambda_k T^*e_k = \lambda_k\overline{\lambda_k}e_k = |\lambda_k|^2 e_k.$$

By the linear map lemma, there exists $R \in \text{End}(V)$ such that $Re_k = |\lambda_k|e_k$ for each k. Then $\text{Mat}(R, (e_1, \ldots, e_n))$ is diagonal with entries $|\lambda_1|, \ldots, |\lambda_n|$, so R is positive by 7.38(c) and (a). Moreover, $R^2e_k = |\lambda_k|^2 e_k = T^*Te_k$ for each k, so $R^2 = T^*T$ by the linear map lemma. Since R is a positive square root of T^*T, we have $R = \sqrt{T^*T}$ by 7.39. By the linear map lemma again, there exists $S \in \text{End}(V)$ such that $Se_k = \mu_k e_k$, where

$$\mu_k = \begin{cases} \frac{\lambda_k}{|\lambda_k|}, & \lambda_k \neq 0, \\ 1, & \lambda_k = 0. \end{cases}$$

Since each $Se_k = \mu_k e_k$ with $|\mu_k| = 1$, it follows that $Se_1, \ldots, Se_n$ is an orthonormal basis of V, so S is unitary by 7.53(d). Moreover, $Te_k = \lambda_k e_k = \mu_k|\lambda_k|e_k = SRe_k$

for each k, so $T = SR = S\sqrt{T^*T}$ by the linear map lemma. It now follows that $\text{Mat}(S, (e_1, \dots, e_n))$ is diagonal with entries $\mu_1, \dots, \mu_n$, and $\text{Mat}(\sqrt{T^*T}, (e_1, \dots, e_n))$ is diagonal with entries $|\lambda_1|, \dots, |\lambda_n|$, so S and $\sqrt{T^*T}$ both have diagonal matrices with respect to the same orthonormal basis $e_1, \dots, e_n$. □

Solution 7F.32 Let $B = \{w \in \text{im}\, T : \|w\| < 1\}$. Since $T \neq 0$, we have $\dim \text{im}\, T = m \geq 1$ by 7.68(b). By the singular value decomposition, there exist orthonormal lists $e_1, \dots, e_m$ in V and $f_1, \dots, f_m$ in W such that

$$Tv = \sum_{k=1}^{m} s_k \langle v, e_k \rangle f_k$$

for all $v \in V$. Moreover, for any $u \in \ker T$,

$$Tu = \sum_{k=1}^{m} s_k \langle u, e_k \rangle f_k = 0.$$

Since $f_1, \dots, f_m$ is linearly independent by 6.25 and $s_k > 0$ for all k, $\langle u, e_k \rangle = 0$ for all $k = 1, \dots, m$, so $e_k \in (\ker T)^\perp$ for all k. By 6.51 and the fundamental theorem of linear maps, $\dim(\ker T)^\perp = \dim \text{im}\, T = m$, so $e_1, \dots, e_m$ form an orthonormal basis of $(\ker T)^\perp$ by 6.28. By 7.96 in $(\ker T)^\perp$,

$$E = E\left(\frac{e_1}{s_1}, \dots, \frac{e_m}{s_m}\right) = \left\{v \in (\ker T)^\perp : \sum_{k=1}^{m} s_k^2 |\langle v, e_k \rangle|^2 < 1\right\}.$$

By 6.24, $\|Tv\|^2 = \sum_{k=1}^{m} s_k^2 |\langle v, e_k \rangle|^2$ for all $v \in V$, so

$$E = \{v \in (\ker T)^\perp : \|Tv\| < 1\}.$$

We now show that $T(E) = B$.

$\subseteq$ This follows immediately from the definition of E.

$\supseteq$ Suppose $w \in B$. By 6.67, there exists $v \in (\ker T)^\perp$ such that $Tv = w$. Then $\|Tv\| < 1$, so $v \in E$, and hence $w = Tv \in T(E)$. □

Chapter 8

§8A

Solution 8A.1 By 8.1, $\ker T^4 \subseteq \ker T^5 \subseteq \ker T^6$. Then $8 \leq \dim \ker T^5 \leq 9$ by 2.37. If $\dim \ker T^5 = 8$, then 2.39 gives $\ker T^4 = \ker T^5$. 8.2 yields $\ker T^4 = \ker T^6$, contradicting $\dim \ker T^6 = 9$. Therefore, $\dim \ker T^5 = 9$, so $\ker T^5 = \ker T^6$ by 2.39. Hence, $\ker T^5 = \ker T^m$ for all $m \geq 5$ by 8.2, so $\dim \ker T^m = 9$ for all $m \geq 5$. □

Solution 8A.2 By 3.10, $T^j v = T^{j-m}(T^m v) = T^{j-m}(0) = 0$ for all $j \geq m$. Suppose

$$\sum_{j=0}^{m-1} \lambda_j T^j v = 0$$

for some scalars $\lambda_0, \ldots, \lambda_{m-1} \in \mathbb{F}$, where not all λ_j are zero. Let k be the smallest index such that $\lambda_k \neq 0$. Applying T^{m-1-k} to both sides gives

$$\sum_{j=0}^{m-1} \lambda_j T^{j+m-1-k} v = 0.$$

If $j < k$, then $\lambda_j = 0$ by the choice of k. If $j > k$, then $j + m - 1 - k \geq m$, so $T^{j+m-1-k} v = 0$ by the observation above. Hence, the sum reduces to $\lambda_k T^{m-1} v = 0$. Since $\lambda_k \neq 0$, it must be that $T^{m-1} v = 0$, which contradicts the assumption that $T^{m-1} v \neq 0$. Therefore, all $\lambda_j = 0$, proving that $v, Tv, \ldots, T^{m-1} v$ is linearly independent. □

Solution 8A.3 By Exercise 5D4, it suffices to show that $\ker T \cap \operatorname{im} T = \{0\}$ if and only if $\ker T^2 = \ker T$. By 8.1, $\ker T \subseteq \ker T^2$. Moreover, $0 \in \ker T \cap \operatorname{im} T$. Consequently,

$$\begin{aligned}
\ker T^2 = \ker T &\iff \ker T^2 \subseteq \ker T \\
&\iff \text{for all } v \in V,\ T^2 v = 0 \text{ implies } Tv = 0 \\
&\iff \text{for all } w \in \operatorname{im} T,\ Tw = 0 \text{ implies } w = 0 \\
&\iff \ker T \cap \operatorname{im} T = \{0\}.
\end{aligned}$$

□

Solution 8A.4 Let p be the minimal polynomial of T. By assumption, $p(z) = (z - \lambda)^m q(z)$ for some $q \in \mathbb{F}[X]$, which is monic since p and $(z - \lambda)^m$ are monic. We claim that for each $k = 1, \ldots, m$,

$$\ker(T - \lambda I)^{k-1} \subsetneq \ker(T - \lambda I)^k.$$

Indeed, if equality did hold for some k, then 8.2 applied to $T - \lambda I$ yields $\ker(T - \lambda I)^{k-1} = \ker(T - \lambda I)^m$. Then for each $v \in V$,

$$(T - \lambda I)^m (q(T)v) = p(T)v = 0$$

by 5.17(a), so $q(T)v \in \ker(T - \lambda I)^m = \ker(T - \lambda I)^{k-1}$. Let $r(z) = (z - \lambda)^{k-1} q(z)$. Then r is monic, and

$$\deg r = \deg p - (m - k + 1) < \deg p.$$

Additionally, $r(T)v = (T - \lambda I)^{k-1} q(T)v = 0$ for all $v \in V$ by 5.17(a), whence $r(T) = 0$. This contradicts the minimality of p by 5.22, so equality cannot hold for any $k = 1, \ldots, m$. Thus,

$$\{0\} = \ker(T - \lambda I)^0 \subsetneq \ker(T - \lambda I)^1 \subsetneq \cdots \subsetneq \ker(T - \lambda I)^m.$$

Since strict inclusion raises the dimension by at least 1 by 2.37 and 2.39, we have

$$\dim \ker (T - \lambda I)^m \geq m.$$

□

Solution 8A.5

Axler Hint. *Exercise 3B21 may be useful.*

We proceed by induction on m. The base case $m = 1$ is trivial since $\dim \ker T^1 = \dim \ker T \leq 1 \cdot \dim \ker T$. Suppose $m \geq 2$ and the statement holds for $m - 1$. Since $v \in \ker T^m$ if and only if $T^{m-1} T v = 0$, i.e., $Tv \in \ker T^{m-1}$, we have

$$\ker T^m = \{v \in V : Tv \in \ker T^{m-1}\}.$$

By Exercise 3B21 with $W = V$ and $U = \ker T^{m-1}$, we have

$$\begin{aligned} \dim \ker T^m &= \dim \ker T + \dim(\ker T^{m-1} \cap \operatorname{im} T) \\ &\leq \dim \ker T + \dim \ker T^{m-1} \\ &\leq \dim \ker T + (m-1) \dim \ker T \\ &= m \dim \ker T, \end{aligned}$$

where the first inequality follows from 2.37, and the second inequality follows from the induction hypothesis. This completes the induction step and proves the statement for all positive integers m. □

Solution 8A.6 Note that $\operatorname{im} T^0 = \operatorname{im} I = V$. For each $k \geq 0$, every $v \in \operatorname{im} T^{k+1}$ has the form $v = T^{k+1} u$ for some $u \in V$, hence $v = T^k(Tu) \in \operatorname{im} T^k$. Therefore, $\operatorname{im} T^{k+1} \subseteq \operatorname{im} T^k$ for all $k \geq 0$. □

Solution 8A.7 Observe that for each $j \geq 0$,

$$\operatorname{im} T^{j+1} = \{T(T^j v) : v \in V\} = \{Tw : w \in \operatorname{im} T^j\}.$$

We proceed by induction on $k \geq m + 1$. The base case $k = m + 1$ is given. Suppose $k \geq m + 1$ and $\operatorname{im} T^k = \operatorname{im} T^m$. Then

$$\operatorname{im} T^{k+1} = \{Tw : w \in \operatorname{im} T^k\} = \{Tw : w \in \operatorname{im} T^m\} = \operatorname{im} T^{m+1} = \operatorname{im} T^m.$$

This completes the induction step and proves the statement for all $k > m$. □

Solution 8A.8 By 8.3, $\ker T^{\dim V + k} = \ker T^{\dim V}$ for all $k \geq 0$. By the fundamental theorem of linear maps,

$$\dim \operatorname{im} T^{\dim V + k} = \dim V - \dim \ker T^{\dim V + k} = \dim V - \dim \ker T^{\dim V} = \dim \operatorname{im} T^{\dim V}$$

for all $k \geq 0$. By Exercise 8A6, $\operatorname{im} T^{\dim V + k} \subseteq \operatorname{im} T^{\dim V}$ for all $k \geq 0$. Equality follows from 2.39. □

Solution 8A.9 By 8.1 and Exercise 8A6, $\ker T^m \subseteq \ker T^{m+1}$ and $\operatorname{im} T^{m+1} \subseteq \operatorname{im} T^m$. By 2.39, each inclusion is an equality if and only if the dimensions are equal. By the fundamental theorem of linear maps, $\dim \ker T^j = \dim V - \dim \operatorname{im} T^j$ for all $j \geq 0$. Therefore,

$$\begin{aligned}
\ker T^m = \ker T^{m+1} &\iff \dim \ker T^m = \dim \ker T^{m+1} \\
&\iff \dim V - \dim \operatorname{im} T^m = \dim V - \dim \operatorname{im} T^{m+1} \\
&\iff \dim \operatorname{im} T^m = \dim \operatorname{im} T^{m+1} \\
&\iff \operatorname{im} T^m = \operatorname{im} T^{m+1}.
\end{aligned}$$

□

Solution 8A.10 Observe that $T^2(w,z) = T(z,0) = (0,0)$, so $T^2 = 0$, and hence T is nilpotent. By 8.17(a), the only eigenvalue of T is 0. Therefore, every generalized eigenvector of T corresponds to 0. For every nonzero $v \in \mathbb{C}^2$, we have $T^2 v = 0$, so v is a generalized eigenvector of T corresponding to 0. □

Solution 8A.11

Axler Hint. *Assume* $\mathbb{F} = \mathbb{R}$ *because the case* $\mathbb{F} = \mathbb{C}$ *follows from 5.27(b) and 8.9.*

If $\mathbb{F} = \mathbb{C}$, the result follows from 5.27(b) and 8.9, so we may assume $\mathbb{F} = \mathbb{R}$.

We first claim that a monic $q_0 \in \mathbb{F}[X]$ that divides $q(z) = (z - \mu_1) \cdots (z - \mu_m)$ for some $\mu_1, \ldots, \mu_m \in \mathbb{F}$ is a product of linear factors, each of the form $z - \mu_k$ for some k. Write $q = q_0 r$, and factor q_0 and r by 4.16. By the uniqueness in 4.16, q_0 consists of linear factors of the form $z - \mu_k$.

Now, let $n = \dim V$ and let p be the minimal polynomial of T.

$\Rightarrow$ Suppose $v_1, \ldots, v_n$ is a basis of V consisting of generalized eigenvectors of T with corresponding eigenvalues $\nu_1, \ldots, \nu_n$. Then there exist positive integers $k_1, \ldots, k_n$ such that $(T - \nu_\ell I)^{k_\ell} v_\ell = 0$ for each $\ell = 1, \ldots, n$. Consider the polynomial

$$s(z) = \prod_{\ell=1}^{n} (z - \nu_\ell)^{k_\ell}.$$

By 5.17, the factors of $s(T) = \prod_\ell (T - \nu_\ell I)^{k_\ell}$ commute with each other. Moving $(T - \nu_\ell I)^{k_\ell}$ to the right gives $s(T) v_\ell = 0$ for each ℓ, so $s(T) = 0$ because $v_1, \ldots, v_n$ is a basis of V. By 5.29, p divides s, so the preliminary claim implies that p is a product of linear factors, each of the form $z - \nu_\ell$ for some ℓ.

$\Leftarrow$ We proceed by strong induction on n. Suppose $p(z) = (z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. Suppose the result holds for all operators on all vector spaces of dimension less than n whose minimal polynomial is a product of linear factors. Let $U = \operatorname{im}(T - \lambda_1 I)^n$, so $V = \ker(T - \lambda_1 I)^n \oplus U$ by 8.4.

If $U = \{0\}$, then $V = \ker(T - \lambda_1 I)^n$ by 8.4, and any basis of V works. Otherwise, U is nontrivial and invariant under T by 5.18. Since λ_1 is an eigenvalue of T by 5.27(a), $\{0\} \neq \ker(T - \lambda_1 I) \subseteq \ker(T - \lambda_1 I)^n$, so $\dim U < n$. By 5.31 and the preliminary claim, the minimal polynomial of $T|_U$ is a product of linear factors. If $u \in U$ is a generalized eigenvector of $T|_U$ corresponding to an eigenvalue λ, then $(T - \lambda I)^k u = (T|_U - \lambda I)^k u = 0$ for some positive integer k. Additionally, λ is an

eigenvalue of T because an eigenvector of $T|_U$ is also an eigenvector of T. Hence, u is a generalized eigenvector of T corresponding to λ. By the induction hypothesis and the preceding argument, U has a basis consisting of generalized eigenvectors of T. Concatenating any basis of $\ker(T-\lambda_1 I)^n$ with the basis of U from above gives a basis of V consisting of generalized eigenvectors of T by 8.4. □

Solution 8A.12 Let $v \in V$ be a nonzero vector, and let $\dim V = n$. By 8.1 and 8.3, $\ker(T-\alpha I)^k \subseteq \ker(T-\alpha I)^n$ for all $\alpha \in \mathbb{F}$ and positive integers k. By hypothesis, v is a generalized eigenvector of T corresponding to some eigenvalue $\lambda \in \mathbb{F}$. We claim that every nonzero vector in V is a generalized eigenvector of T corresponding to the same eigenvalue λ.

Suppose some nonzero $u \in V$ is not a generalized eigenvector of T corresponding to λ. By hypothesis and 8.11, u corresponds to a unique eigenvalue $\mu \in \mathbb{F}$, where $\mu \neq \lambda$ by choice of u. By 8.12, the list v, u is linearly independent, so $v+u$ is a nonzero vector in V. By hypothesis and 8.11 again, $v+u$ is a generalized eigenvector of T corresponding to a unique eigenvalue $\nu \in \mathbb{F}$. If $\nu = \lambda$, then $u = (v+u) - v \in \ker(T-\lambda I)^n$, which contradicts 8.11, since u corresponds to $\mu \neq \lambda$. If $\nu = \mu$, then $v = (v+u) - u \in \ker(T-\mu I)^n$, which contradicts 8.11, since v corresponds to $\lambda \neq \mu$. Thus, $\nu \neq \lambda$ and $\nu \neq \mu$, whence the list $v, u, v+u$ is linearly independent by 8.12, which is a contradiction since $v+u \in \operatorname{span}(v,u)$. Therefore, every nonzero vector in V is a generalized eigenvector of T corresponding to the same eigenvalue λ. Hence, each nonzero vector of V lies in $\ker(T-\lambda I)^k$ for some positive integer k, and thus lies in $\ker(T-\lambda I)^n$. Moreover, $(T-\lambda I)^n 0 = 0$, so $(T-\lambda I)^n = 0$, which shows that $T-\lambda I$ is nilpotent. □

Solution 8A.13 Since ST is nilpotent, there exists a positive integer k such that $(ST)^k = 0$. Then

$$(TS)^{k+1} = T(ST)^k S = T0S = 0,$$

so TS is nilpotent. □

Solution 8A.14 Suppose T is diagonalizable. By 8.17(a), the only eigenvalue of T is 0. Then 5.55(c) gives $V = E(0,T) = \ker T$, so $T = 0$, which contradicts the assumption that $T \neq 0$. □

Solution 8A.15

Axler Note. *For $\mathbb{F} = \mathbb{C}$, this exercise adds another equivalence to the list of conditions for diagonalizability in 5.55.*

$\Rightarrow$ Suppose T is diagonalizable, and let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. By 5.55(c), we have

$$V = \bigoplus_{k=1}^{m} E(\lambda_k, T).$$

Let $v \in V$ be a generalized eigenvector of T corresponding to an eigenvalue λ, so $\lambda = \lambda_j$ for some j, and $(T-\lambda I)^d v = 0$ for some positive integer d. Write $v = \sum_{k=1}^{m} v_k$

with $v_k \in E(\lambda_k, T)$. Then

$$\sum_{k=1}^{m}(T-\lambda I)^d v_k = \sum_{k=1}^{m}(\lambda_k-\lambda)^d v_k = 0.$$

Because the sum is direct, $(\lambda_k-\lambda)^d v_k = 0$ for each k. For $\lambda_k \neq \lambda$, this implies $v_k = 0$, so $v = v_j \in E(\lambda, T)$, which shows that v is an eigenvector of T corresponding to λ.

$\Leftarrow$ Suppose every generalized eigenvector of T is an eigenvector of T. By 8.9, V has a basis consisting of generalized eigenvectors of T, which are eigenvectors of T by assumption, so 5.55(b) implies that T is diagonalizable. □

Solution 8A.16

(a) Let $V = \mathbb{F}^2$ and define $S, T \in \text{End}(V)$ by $S(x, y) = (y, 0)$ and $T(x, y) = (0, x)$. Then $S^2(x, y) = S(y, 0) = (0, 0)$ and $T^2(x, y) = T(0, x) = (0, 0)$, so $S^2 = T^2 = 0$, which shows that S and T are nilpotent. Moreover,

$$(S + T)(1, 1) = (1, 1), \quad ST(1, 0) = (1, 0),$$

from $(S + T)(x, y) = (y, x)$ and $ST(x, y) = (x, 0)$. Then $(1, 1)$ and $(1, 0)$ are eigenvectors of $S + T$ and ST, respectively, corresponding to the eigenvalue 1, so neither $S + T$ nor ST is nilpotent by 8.17(a).

(b) By 8.16, $S^n = T^n = 0$, where $n = \dim V$. Since S and T commute, $(ST)^n = S^n T^n = 0$, so ST is nilpotent. By the commutativity of S and T and the distributive property of linear maps,

$$(S + T)^{2n-1} = \sum_{k=0}^{2n-1} \binom{2n-1}{k} S^k T^{2n-1-k}.$$

For $k \geq n$, $S^k = S^n S^{k-n} = 0$, and for $2n-1-k \geq n$, $T^{2n-1-k} = T^n T^{n-1-k} = 0$. If $k \leq n-1$, then $2n-1-k \geq n$, so at least one of S^k or T^{2n-1-k} is 0. Hence, every term in the sum is 0, so $(S + T)^{2n-1} = 0$, which shows that $S + T$ is nilpotent.

□

Solution 8A.17

(a) Let $R = I + T + \cdots + T^{m-1}$. Then

$$(I - T)R = (I + T + \cdots + T^{m-1}) - (T + T^2 + \cdots + T^m) = I - T^m = I.$$

By 5.17(b), $(I - T)R = R(I - T) = I$, so $I - T$ is invertible with inverse R.

(b) Recall the geometric series formula

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k, \quad |x| < 1.$$

Replacing x with T and 1 with I, this suggests that

$$(I - T)^{-1} = \sum_{k=0}^{\infty} T^k.$$

Since $T^k = T^{k-m}T^m = 0$ for $k \geq m$, the series terminates at $k = m-1$, yielding

$$(I - T)^{-1} = I + T + \cdots + T^{m-1}. \qquad \square$$

Solution 8A.18

Axler Note. *If* $\dim \operatorname{im} T < \dim V - 1$, *then this exercise improves 8.16.*

Given that T is nilpotent, there exists a positive integer k such that $T^k = 0$. Since $\operatorname{im} T$ is invariant under T, we can consider the restriction $S = T|_{\operatorname{im} T} \in \operatorname{End}(\operatorname{im} T)$. Because $S^k = T^k|_{\operatorname{im} T} = 0$, S is nilpotent. Applying 8.16 to S, we get $S^{\dim \operatorname{im} T} = 0$. Then for every $v \in V$,

$$T^{1+\dim \operatorname{im} T} v = S^{\dim \operatorname{im} T}(Tv) = 0. \qquad \square$$

Solution 8A.19 If the first n inclusions in the chain of 8.1 were strict, then

$$\{0\} = \ker T^0 \subsetneq \ker T^1 \subsetneq \cdots \subsetneq \ker T^n,$$

where the dimension of each successive null space increases by at least 1, so $\dim \ker T^n \geq n = \dim V$. By 2.39, $\ker T^n = V$, so $T^n = 0$, contradicting that T is not nilpotent. Hence, there must be some $k \leq n-1$ such that $\ker T^k = \ker T^{k+1}$, so by 8.2, $\ker T^{n-1} = \ker T^n$. Since $\operatorname{im} T^n \subseteq \operatorname{im} T^{n-1}$ and $\dim \operatorname{im} T^n = \dim \operatorname{im} T^{n-1}$ by the fundamental theorem of linear maps, it follows that $\operatorname{im} T^{n-1} = \operatorname{im} T^n$ by 2.39. By 8.4,

$$V = \ker T^n \oplus \operatorname{im} T^n = \ker T^{n-1} \oplus \operatorname{im} T^{n-1}. \qquad \square$$

Solution 8A.20

Appendix. *An alternate proof is given at the end of this solution.*

Let $n = \dim V$. By 7.21(c), $V = \ker T \oplus \operatorname{im} T$. By Exercise 8A3, $\ker T^2 = \ker T$. Applying 8.2 yields $\ker T^k = \ker T$ for all $k \geq 1$. By 8.16, $T^n = 0$, so $V = \ker T^n = \ker T$, and hence $T = 0$.

Alternate proof. Civil Service Pigeon. If $T \neq 0$, let $k \geq 2$ be the smallest integer such that $T^k = 0$. Then there exists $v \in V$ with $T^{k-1}v \neq 0$. Then $T(T^{k-1}v) = T^k v = 0$, meaning $T^{k-1}v \in \operatorname{null} T$. Since T is normal, $\operatorname{null} T = \operatorname{null} T^*$, so $T^*(T^{k-1}v) = 0$. Thus

$$\|T^{k-1}v\|^2 = \langle T^{k-1}v, T^{k-1}v\rangle = \langle T^{k-2}v, T^*T^{k-1}v\rangle = 0 \implies T^{k-1}v = 0,$$

contradicting the choice of k. Therefore, $T = 0$. $\square$

Solution 8A.21 Let $n = \dim V$. If $\ker T^k = \ker T^{k+1}$ for some $k \in \{0, \dots, n-1\}$, then 8.2 gives $\ker T^{n-1} = \ker T^n$, contradicting the hypothesis. Hence, every inclusion in the chain in 8.1 must be strict:

$$\{0\} = \ker T^0 \subsetneq \ker T^1 \subsetneq \cdots \subsetneq \ker T^n.$$

It follows that the dimension of each successive null space increases by at least 1. Therefore, for $0 \le k \le n$,

$$\dim \ker T^k \ge k \quad \text{and} \quad n \ge \dim \ker T^n \ge \dim \ker T^k + (n-k).$$

This forces $\dim \ker T^k = k$ for all $0 \le k \le n$. By 2.39, $\ker T^n = V$, so $T^n = 0$, and hence T is nilpotent. □

Solution 8A.22 By Exercise 8A9 with $m = 4$, $\ker T^4 \neq \ker T^5$. Since $\dim(\mathbb{C}^5) = 5$, Exercise 8A21 yields that T is nilpotent. □

Solution 8A.23

Axler Note. *This exercise shows that (b) in 8.17 does not hold without the hypothesis that* $\mathbb{F} = \mathbb{C}$.

Define $T \in \text{End}(\mathbb{R}^3)$ by $T(x, y, z) = (0, -z, y)$. Since $T(1, 0, 0) = (0, 0, 0)$, it follows that 0 is an eigenvalue of T. Suppose (x, y, z) is an eigenvector of T with eigenvalue λ. Then $T(x, y, z) = \lambda(x, y, z)$ gives

$$\begin{cases} \lambda x = 0, \\ \lambda y = -z, \\ \lambda z = y. \end{cases}$$

The two equations $\lambda y = -z$ and $\lambda z = y$ imply $(\lambda^2 + 1)y = 0$. Since $\lambda \in \mathbb{R}$, we have $\lambda^2 + 1 \neq 0$, so $y = 0$, whence $z = -\lambda y = 0$. Since (x, y, z) is an eigenvector, it is nonzero, so $x \neq 0$, and the first equation $\lambda x = 0$ forces $\lambda = 0$. Thus, 0 is the only eigenvalue of T. Moreover, $T^3(0, 1, 0) = T^2(0, 0, 1) = T(0, -1, 0) = (0, 0, -1) \neq (0, 0, 0)$, so $T^3 \neq 0$, and hence T is not nilpotent by 8.16. □

Solution 8A.24

(a) Let $e_1, \dots, e_4$ be the standard basis of $\mathbb{F}^4$. Then $Te_1 = e_3$ and $Te_2 = e_4$, while $Te_3 = Te_4 = 0$, so a basis of $\ker T$ is e_3, e_4. Since $\ker T^2 = \mathbb{F}^4$ by the example, we can extend the basis of $\ker T$ to a basis e_3, e_4, e_1, e_2 of $\mathbb{F}^4$. Therefore,

$$\text{Mat}(T, (e_3, e_4, e_1, e_2)) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

(b) Let $T \in \text{End}(\mathbb{F}^3)$ be the operator whose matrix with respect to the standard basis is the matrix A given in the example, and consider $v = (0, 1, 0)$. Then $Tv = (9, 9, 0)$, the second column of A, and $T^2 v = T(9, 9, 0) = (54, 18, 36) \neq 0$.

Lastly, $T^3v = 0$ since $A^3 = 0$ as given in the example. By Exercise 8A2, the vectors v, Tv, T^2v are linearly independent, and hence form a basis of $\mathbb{F}^3$ by 2.38. Therefore,

$$\operatorname{Mat}(T, (T^2v, Tv, v)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

(c) Consider the standard basis of $\mathbb{R}[X]_{\leq m}$, and let D be the differentiation operator. Then $D1 = 0$ and $Dx^k = kx^{k-1}$ for $k = 1, \ldots, m$. Therefore, the matrix of D with respect to the standard basis has entries $1, 2, \ldots, m$ directly above the diagonal and 0 elsewhere.

□

Solution 8A.25 By 8.18(b), the minimal polynomial of T is $z^m = (z-0)^m$ for some positive integer m. By 6.37, the matrix of T is upper triangular with respect to an orthonormal basis of V. By 5.41, the diagonal entries of the matrix are the eigenvalues of T, hence all 0 by 8.17(a). □

§8B

Solution 8B.1 If $T(w, z) = \lambda(w, z)$ for $\lambda \in \mathbb{C}$ and nonzero $(w, z) \in \mathbb{C}^2$, then $-z = \lambda w$ and $w = \lambda z$. Together, these equations imply $(\lambda^2 + 1)z = 0$. Since $z \neq 0$, for otherwise $w = 0$, we must have $\lambda^2 + 1 = 0$, whence $\lambda = \pm i$. If $\lambda = i$, then $-z = iw$, so $E(i, T) = \operatorname{span}((i, 1))$. A similar argument shows that $E(-i, T) = \operatorname{span}((-i, 1))$. Since $E(\pm i, T) \subseteq G(\pm i, T)$, each generalized eigenspace must have dimension at least 1. Moreover, the multiplicities must sum to 2 by 8.25, so each generalized eigenspace has dimension exactly 1. By 2.39,

$$G(i, T) = \operatorname{span}((i, 1)), \quad G(-i, T) = \operatorname{span}((-i, 1)).$$ □

Solution 8B.2 Observe that $T^{-1} - \frac{1}{\lambda}I = -\frac{1}{\lambda}T^{-1}(T - \lambda I)$. Moreover, $T^{-1}(T - \lambda I) = (T - \lambda I)T^{-1} = I - \lambda T^{-1}$. Then for all positive integers k,

$$\left(T^{-1} - \frac{1}{\lambda}I\right)^k = \left(-\frac{1}{\lambda}T^{-1}(T - \lambda I)\right)^k = \left(-\frac{1}{\lambda}\right)^k T^{-k}(T - \lambda I)^k.$$

Since $\lambda \neq 0$, the operator $(-\frac{1}{\lambda})^k T^{-k}$ is injective, and therefore $\ker(T^{-1} - \frac{1}{\lambda}I)^k = \ker(T - \lambda I)^k$. By 8.19, the result follows. □

Solution 8B.3 Note that

$$(S^{-1}TS - \lambda I)^k = (S^{-1}(T - \lambda I)S)^k = S^{-1}(T - \lambda I)^k S.$$

Because S^{-1} is injective, $S^{-1}(T - \lambda I)^k Sv = 0$ if and only if $(T - \lambda I)^k Sv = 0$, and therefore $\ker(S^{-1}TS - \lambda I)^k = S^{-1}(\ker(T - \lambda I)^k)$ for all positive integers k. For $k = 1$,

we have $E(\lambda, S^{-1}TS) = S^{-1}(E(\lambda, T))$. Since S^{-1} is injective, $E(\lambda, S^{-1}TS) \neq \{0\}$ if and only if $E(\lambda, T) \neq \{0\}$, so T and $S^{-1}TS$ have the same eigenvalues. Taking $k = \dim V$ and using 8.20 yields $G(\lambda, S^{-1}TS) = S^{-1}(G(\lambda, T))$. Note that $S^{-1}|_{G(\lambda,T)} \in \operatorname{Hom}(G(\lambda, T), V)$ is an injective linear map whose range is $G(\lambda, S^{-1}TS)$. By the fundamental theorem of linear maps,

$$\dim G(\lambda, T) = \dim \ker S^{-1}|_{G(\lambda,T)} + \dim \operatorname{im} S^{-1}|_{G(\lambda,T)} = \dim G(\lambda, S^{-1}TS),$$

so each eigenvalue has the same multiplicity for T as for $S^{-1}TS$. □

Solution 8B.4 Let $n = \dim V$. Because $\ker T^{n-2} \neq \ker T^{n-1}$, the contrapositive of 8.2 gives $\ker T^k \neq \ker T^{k+1}$ for all $k \leq n-2$. Then all the inclusions in the chain up to $\ker T^{n-1}$ in 8.1 must be strict:

$$\{0\} = \ker T^0 \subsetneq \ker T^1 \subsetneq \cdots \subsetneq \ker T^{n-1}.$$

Consequently, the dimension of each successive null space increases by at least one, so $\dim \ker T^{n-1} \geq n-1$. Because $\ker T \neq \{0\}$, 5.7 gives that 0 is an eigenvalue of T. By 8.1 and 8.20, $\ker T^{n-1} \subseteq \ker T^n = G(0, T)$, so $\dim G(0, T) \geq \dim \ker T^{n-1} \geq n-1$. Let $0, \lambda_2, \ldots, \lambda_m$ be the distinct eigenvalues of T. By 8.12 and 1.45, $G(0, T) + G(\lambda_2, T) + \cdots + G(\lambda_m, T)$ is a direct sum, so by 3.94 and 2.37,

$$\dim G(0, T) + \dim G(\lambda_2, T) + \cdots + \dim G(\lambda_m, T) \leq \dim V = n.$$

Since each dimension must be at least 1, $(n-1) + (m-1) \leq n$, so $m \leq 2$. □

Solution 8B.5

Manual Remark. *In fact, if T has m distinct nonzero eigenvalues, then $V = \ker T^{n-m} \oplus \operatorname{im} T^{n-m}$.*

By 5.12, $n \geq 2$ since T has at least two distinct eigenvalues, 3 and 8. We claim that $\ker T^{n-2} = \ker T^{n-1}$. If 0 is not an eigenvalue of T, then 5.7 yields $\ker T^0 = \ker T^1 = \{0\}$, so 8.2 gives $\ker T^{n-2} = \ker T^{n-1}$. If 0 is an eigenvalue of T, then $0, 3, 8$ are three distinct eigenvalues, and the contrapositive of Exercise 8B4 gives $\ker T^{n-2} = \ker T^{n-1}$. In either case, we have $\ker T^{n-2} = \ker T^n$ by 8.2. Applying Exercise 8A9 with $m = n-2$ yields $\operatorname{im} T^{n-2} = \operatorname{im} T^{n-1}$, so $\operatorname{im} T^{n-2} = \operatorname{im} T^n$ by Exercise 8A7. Therefore, 8.4 gives

$$V = \ker T^n \oplus \operatorname{im} T^n = \ker T^{n-2} \oplus \operatorname{im} T^{n-2}.$$ □

Solution 8B.6 Let $n = \dim V$. Let p be the minimal polynomial of T. By 5.27(a), λ is a zero of p, so applying 4.6 repeatedly gives $p(z) = (z-\lambda)^k q(z)$ for some positive integer k and some $q \in \mathbb{F}[X]$ with $q(\lambda) \neq 0$. Let $G = G(\lambda, T)$ and $U = \operatorname{im}(T - \lambda I)^n$. By 8.4, $V = G \oplus U$, and both G and U are invariant under T by 5.18. Because $(T-\lambda I)^n|_G = 0$ by 8.20, there exists a smallest positive integer m such that $(T - \lambda I)^m|_G = 0$. We claim that $m = k$.

$\leq$ By 5.29, the minimal polynomial r of $T|_G$ divides $(z-\lambda)^m$, so by 4.13 if $\mathbb{F} = \mathbb{C}$ or 4.16 if $\mathbb{F} = \mathbb{R}$, we have $r(z) = (z-\lambda)^{m'}$ for some $m' \leq m$. Because λ is an eigenvalue

of T, we have $\{0\} \neq E(\lambda, T) \subseteq G$. It follows that $r \neq 1$, so $m' \geq 1$. The minimality of m gives $m' = m$, so the minimal polynomial of $T|_G$ is $(z-\lambda)^m$. By 5.31, r also divides p. If $m > k$, then $(z-\lambda)^{m-k}$ would divide q, contradicting $q(\lambda) \neq 0$. Therefore, $m \leq k$.

$\geq$ Note that $\ker(T-\lambda I) \subseteq G$ and $G \cap U = \{0\}$, so $(T-\lambda I)|_U$ is injective. For any $u \in U$, we have $(T-\lambda I)^k q(T)u = 0$, where $q(T)u \in U$. By injectivity, this implies $q(T)u = 0$. Let $s(z) = (z-\lambda)^m q(z)$. Then $s(T)u = (T-\lambda I)^m q(T)u = (T-\lambda I)^m 0 = 0$. On G, we have $(T-\lambda I)^m|_G = 0$, so $s(T)|_G = q(T)(T-\lambda I)^m|_G = 0$ as well. Therefore, $s(T) = 0$ on all of V. By 5.29, p divides s, so $m \geq k$, whence $m = k$. □

Solution 8B.7

Axler Note. *If $d < \dim V$, then this exercise improves 8.20.*

Let $n = \dim V$ and let $G = G(\lambda, T)$. Define $N = (T-\lambda I)|_G$. By 8.23, $\dim G = d$.

$\subseteq$ By 8.20, $G = \ker(T-\lambda I)^n$, which is invariant under T by 5.18 applied to $(z-\lambda)^n$, and $N^n = 0$. Thus, $N \in \mathrm{End}(G)$, so $N^d = 0$ by 8.16. Hence, $G \subseteq \ker(T-\lambda I)^d$.

$\supseteq$ Since λ is an eigenvalue of T, we have $\{0\} \neq E(\lambda, T) \subseteq G$, so $d \geq 1$. Hence, $\ker(T-\lambda I)^d \subseteq G$ by 8.19. Therefore, $G = \ker(T-\lambda I)^d$. □

Solution 8B.8

Axler Note. *The case $\mathbb{F} = \mathbb{C}$ follows immediately from 5.27(b) and the generalized eigenspace decomposition (8.22); thus this exercise is interesting only when $\mathbb{F} = \mathbb{R}$.*

$\Rightarrow$ By 3.94 and 2.42, taking bases of all the $G(\lambda_j, T)$ gives a basis of V consisting of generalized eigenvectors of T. By Exercise 8A11, the minimal polynomial p of T is $p(z) = (z-\mu_1)\cdots(z-\mu_M)$ for $\mu_1, \ldots, \mu_M \in \mathbb{F}$. By 5.27(a), $\{\mu_1, \ldots, \mu_M\} = \{\lambda_1, \ldots, \lambda_m\}$, so each λ_j appears with some positive exponent k_j. Grouping together the factors corresponding to the same λ_j gives the desired form $(z-\lambda_1)^{k_1}\cdots(z-\lambda_m)^{k_m}$.

$\Leftarrow$ By Exercise 8A11, V has a basis of generalized eigenvectors of T. Since each generalized eigenvector belongs to some $G(\lambda_j, T)$, the span of the union of the $G(\lambda_j, T)$ equals V, which gives

$$V = \sum_{j=1}^{m} G(\lambda_j, T).$$

By 8.12 and 1.45, this sum is direct. □

Solution 8B.9 Let $n = \dim V$ and let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. By the generalized eigenspace decomposition, $V = G_1 \oplus \cdots \oplus G_m$, where $G_j = G(\lambda_j, T)$. Define D by $D(v_1 + \cdots + v_m) = \lambda_1 v_1 + \cdots + \lambda_m v_m$ for $v_k \in G_k$. Since the sum is direct, D is well-defined and linear. Define $N = T - D$. Since concatenating bases of all the G_j gives a basis of eigenvectors of D by 3.94 and 2.42, it follows that D is diagonalizable by 5.55.

Suppose $u \in G_k$ for some k. By 8.22(a), G_k is invariant under T. Then $Nu = (T-D)u = Tu - \lambda_k u \in G_k$. It follows that $N^n u = (T-\lambda_k I)^n u = 0$ by 8.20, so $N^n = 0$ on each G_k. Moreover, $Tu \in G_k$, so $DTu = \lambda_k Tu = T(\lambda_k u) = TDu$, which gives $DT = TD$ on each G_k. This gives $DN = D(T-D) = DT - D^2 = TD - D^2 =$

$(T - D)D = ND$ on each G_k. Since V is the direct sum of the G_k, it follows that N is nilpotent and $DN = ND$ on V. □

Solution 8B.10 By Schur's theorem, there exists an orthonormal basis $f_1, \dots, f_n$ of V such that the matrix A of T with respect to this basis is upper triangular. By 5.41, 8.31, and 8.25, the diagonal entries $A_{1,1}, \dots, A_{n,n}$ are the eigenvalues $\lambda_1, \dots, \lambda_n$ of T rearranged in some order. Observe that $Tf_k = \sum_{j \le k} A_{j,k} f_j$ for each k, so

$$\|Tf_k\|^2 = \sum_{j \le k} |A_{j,k}|^2 \ge |A_{k,k}|^2$$

by 6.24. By Exercise 7A5 applied to both bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$, we have

$$\sum_{m=1}^{n} \|Te_m\|^2 = \sum_{m=1}^{n} \|Tf_m\|^2 \ge \sum_{m=1}^{n} |A_{m,m}|^2 = \sum_{m=1}^{n} |\lambda_m|^2. \qquad \square$$

Solution 8B.11 Define $T \in \text{End}(\mathbb{C}^4)$ by $T(z_1, z_2, z_3, z_4) = (7z_1, 7z_2, 8z_3, 8z_4)$. Then $\text{Mat}(T)$ with respect to the standard basis is diagonal with diagonal entries $7, 7, 8, 8$. By 5.41, the eigenvalues of T are $7, 8$, each with multiplicity 2 by 8.31. Hence, the characteristic polynomial of T is $(z - 7)^2(z - 8)^2$. □

Solution 8B.12 Let e_1, e_2, e_3, e_4 be the standard basis of $\mathbb{C}^4$. Consider the operator $T \in \text{End}(\mathbb{C}^4)$ defined by $T(z_1, z_2, z_3, z_4) = (z_1, 5z_2 + z_3, 5z_3, 5z_4)$. Then the matrix of T with respect to the standard basis is

$$\text{Mat}(T) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

By 5.41, the eigenvalues of T are $1, 5$, with 1 having multiplicity 1 and 5 having multiplicity 3 by 8.31. Therefore, the characteristic polynomial of T is $(z - 1)(z - 5)^3$. Let p be the minimal polynomial of T. By 8.30, 5.27(a), and uniqueness of the factorization of the minimal polynomial, we have $p(z) = (z - 1)(z - 5)^b$ for some $b \in \{1, 2, 3\}$. Observe that $(T - I)(T - 5I)e_3 = 4e_2 \ne 0$. However, $(T - 5I)^2 e_1 = 16e_1$ and $(T - 5I)^2 e_k = 0$ for $k = 2, 3, 4$. Moreover, $(T - I)e_1 = 0$, so $(T - I)(T - 5I)^2 = 0$. Hence, $b = 2$ and the minimal polynomial of T is $p(z) = (z - 1)(z - 5)^2$. □

Solution 8B.13 Let e_1, e_2, e_3, e_4 be the standard basis of $\mathbb{C}^4$. Consider the operator $T \in \text{End}(\mathbb{C}^4)$ defined by $T(z_1, z_2, z_3, z_4) = (0, z_2 + z_3, z_3, 3z_4)$. Then the matrix of T with respect to the standard basis is

$$\text{Mat}(T) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

By 5.41, the eigenvalues of T are $0, 1, 3$, with 0 and 3 having multiplicity 1 and 1 having multiplicity 2 by 8.31. Therefore, the characteristic polynomial of T is $z(z-1)^2(z-3)$. Let p be the minimal polynomial of T. By 8.30, 5.27(a), and uniqueness of the factorization of the minimal polynomial, we have $p(z) = z(z-1)^b(z-3)$ for some $b \in \{1, 2\}$. Since $T(T-I)(T-3I)e_3 = -2e_2 \neq 0$, we have $b = 2$, so $p(z) = z(z-1)^2(z-3)$. □

Solution 8B.14 Let $T \in \text{End}(\mathbb{C}^4)$ be defined by $T(z_1, z_2, z_3, z_4) = (0, z_2, z_3, 3z_4)$. Then $\text{Mat}(T)$ with respect to the standard basis is diagonal with diagonal entries $0, 1, 1, 3$. By 5.41, the eigenvalues of T are $0, 1, 3$, with 0 and 3 having multiplicity 1 and 1 having multiplicity 2 by 8.31. Therefore, the characteristic polynomial of T is $z(z-1)^2(z-3)$. Since $\text{Mat}(T)$ is diagonal, T is diagonalizable, so the minimal polynomial of T has distinct linear factors by 5.62, and hence the minimal polynomial of T is $z(z-1)(z-3)$ by 5.27(a). □

Solution 8B.15 Let e_1, e_2, e_3, e_4 be the standard basis of $\mathbb{C}^4$. Observe that $Te_k = e_{k+1}$ for $k = 1, 2, 3$ and $Te_4 = 0$, so $T^4 = 0$, and hence T is nilpotent. By 8.17(a), 0 is the only eigenvalue of T, and $\dim \ker T^4 = \dim \mathbb{C}^4 = 4$, so the multiplicity of 0 is 4 by 8.23, and hence the characteristic polynomial of T is z^4. Moreover, 8.18(b) implies that the minimal polynomial of T is z^m for some positive integer m. By 5.29, $m \leq 4$, and since $T^3e_1 = e_4 \neq 0$, we must have $m \geq 4$, so $m = 4$, and hence the minimal polynomial of T is z^4. □

Solution 8B.16 Let $e_1, \ldots, e_6$ be the standard basis of $\mathbb{C}^6$. Observe that $Te_k = e_{k+1}$ for $k = 1, 2, 4$ and $Te_j = 0$ for $j = 3, 5, 6$. Hence, $T^3 = 0$, and T is nilpotent. By 8.17(a), 0 is the only eigenvalue of T, and $\dim \ker T^6 = \dim \mathbb{C}^6 = 6$, so the multiplicity of 0 is 6 by 8.23, and hence the characteristic polynomial of T is z^6. Moreover, 8.18(b) implies that the minimal polynomial of T is z^m for some positive integer m. By 5.29, $m \leq 3$, and since $T^2e_1 = e_3 \neq 0$, we must have $m \geq 3$, so $m = 3$, and hence the minimal polynomial of T is z^3. □

Solution 8B.17 Let $d = \dim V$. By induction, $P^d = P$, so $V = \ker P^d \oplus \text{im } P^d = \ker P \oplus \text{im } P$ by 8.4. Choose a basis of V consisting of a basis of $\ker P$ followed by a basis of $\text{im } P$. For any $u \in \text{im } P$, there exists some $w \in V$ such that $Pw = u$. Then $Pu = P(Pw) = P^2w = Pw = u$, and hence the matrix of P with respect to this basis is diagonal with m zeros followed by n ones. By 5.41, the eigenvalues of P are the diagonal entries, and the multiplicity of each eigenvalue is the number of times it appears on the diagonal by 8.31. Hence, the characteristic polynomial of P is $z^m(z-1)^n$. □

Solution 8B.18 Let $n = \dim V$. Define $S = T - \lambda I$. By 8.20, $G = G(\lambda, T) = \ker S^n$. Let the integers in each part be denoted by a, b, c, d respectively. The integer a exists because of 5.27(a); b exists because $S^n|_G = 0$; c exists by 8.3; d exists by Exercise 8A9. By Exercise 8B6, $a = b$. Applying Exercise 8A9 to S gives $c = d$. We close the proof by showing $b = c$.

$\leq$ By 8.1 and 8.2, $G \subseteq \ker S^n \subseteq \ker S^{c+n} = \ker S^c$, so $b \leq c$.

$\geq$ By 8.1, $G \subseteq \ker S^b \subseteq \ker S^{b+1} \subseteq G$, so $c \leq b$. Therefore, $b = c$, and hence all four numbers are equal. □

Solution 8B.19 Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of S and let $m_j = \dim G(\lambda_j, S)$ for each j. The characteristic polynomial of S has constant term $(-\lambda_1)^{m_1} \cdots (-\lambda_k)^{m_k}$. Since S is unitary, $|\lambda_j| = 1$ for each j by 7.54, so the absolute value of the constant term is $|(-\lambda_1)^{m_1} \cdots (-\lambda_k)^{m_k}| = 1$. □

Solution 8B.20

Appendix. *An alternate proof is given at the end of this solution.*

Let q be the characteristic polynomial of T. For each k, let $n_k = \dim V_k$. Since each V_k is nonzero and complex, 5.47 yields a basis $v_{k,1}, \ldots, v_{k,n_k}$ of V_k with respect to which $\operatorname{Mat}(T|_{V_k})$ is upper triangular. For each V_k, denote the diagonal entries of $\operatorname{Mat}(T|_{V_k})$ by $d_{k,j}$ for $j = 1, \ldots, n_k$. Let $\mathscr{B}$ be the list obtained by concatenating the bases of the V_k. By 3.94, the sum of n_k over all k equals $\dim V$. Additionally, $\mathscr{B}$ spans V, so it is a basis of V by 2.42. Because $\operatorname{Mat}(T|_{V_k})$ is upper triangular for each k, 5.39(c) yields $Tv_{k,j} \in \operatorname{span}(v_{k,1}, \ldots, v_{k,j})$ for each $j = 1, \ldots, n_k$. Moreover, $\operatorname{span}(v_{k,1}, \ldots, v_{k,j}) \subseteq \operatorname{span}(v_{1,1}, \ldots, v_{1,n_1}, \ldots, v_{k,1}, \ldots, v_{k,j})$, so $\operatorname{Mat}(T, \mathscr{B})$ is upper triangular by 5.39(c), where the diagonal entries are precisely the $d_{k,j}$. By 5.41, 8.31, and 8.26 applied to T and $T|_{V_k}$ for each k, we obtain

$$q = \prod_{k=1}^{m} \prod_{j=1}^{n_k} (z - d_{k,j}) = \prod_{k=1}^{m} p_k. \qquad \square$$

Alternate proof. Civil Service Pigeon. For each eigenvalue $\lambda \in \mathbb{C}$, since $V = \bigoplus_{k=1}^{m} V_k$ with each V_k invariant under T, we have $G(\lambda, T) = \bigoplus_{k=1}^{m} G\left(\lambda, T|_{V_k}\right)$. This implies

$$\dim G(\lambda, T) = \sum_{k=1}^{m} \dim G\left(\lambda, T|_{V_k}\right).$$

The characteristic polynomial of T is

$$\prod_{\lambda} (z-\lambda)^{\dim G(\lambda,T)} = \prod_{\lambda} (z-\lambda)^{\sum_{k=1}^{m} \dim G\left(\lambda,T|_{V_k}\right)} = \prod_{k=1}^{m} \prod_{\lambda} (z-\lambda)^{\dim G\left(\lambda,T|_{V_k}\right)} = p_1 \cdots p_m.$$

Solution 8B.21 Since $\mathbb{C}^{\deg q} \neq \{0\}$, we have $\deg q \geq 1$, so q has a zero by 4.12. Since p and q have the same zeros, p also has a zero. Write $q = pr$ for some $r \in \mathbb{C}[X]$, and factor r by the fundamental theorem of algebra. By uniqueness of factorization in $\mathbb{C}[X]$, we can write

$$q(z) = \prod_{k=1}^{n} (z - \lambda_k)^{m_k}, \quad p(z) = \prod_{k=1}^{n} (z - \lambda_k)^{\ell_k},$$

where $\lambda_1, \ldots, \lambda_n$ are the distinct zeros of q. Moreover, all m_k and ℓ_k are positive integers since p and q have zeros at λ_k, and we have $\ell_k \leq m_k$ for each k since q is a polynomial multiple of p. Note that $\deg q = \sum_{k=1}^{n} m_k$. Let $e_1, \ldots, e_{\deg q}$ be the standard basis of $\mathbb{C}^{\deg q}$, and split it into blocks of sizes $m_1, \ldots, m_n$ corresponding to the distinct zeros $\lambda_1, \ldots, \lambda_n$ of q. For each λ_k, let $v_{k,1}, \ldots, v_{k,m_k}$ be the ordered subset of the standard basis of $\mathbb{C}^{\deg q}$ as described above. By the linear map lemma, there exists a unique linear map $T \in \operatorname{End}(\mathbb{C}^{\deg q})$ such that

$$T v_{k,j} = \begin{cases} \lambda_k v_{k,j} + v_{k,j-1}, & \text{for } j = 2, \ldots, \ell_k, \\ \lambda_k v_{k,j}, & \text{for } j = 1 \text{ and } j = \ell_k + 1, \ldots, m_k. \end{cases}$$

Observe that the matrix of T with respect to the standard basis is upper triangular, where λ_k appears on the diagonal m_k times for each k. Therefore, the characteristic polynomial of T is q by 5.41 and 8.31.

Consider the subspaces $V_k = \operatorname{span}(v_{k,1}, \ldots, v_{k,m_k})$ for each $k = 1, \ldots, n$. Then $(T - \lambda_k I)^{\ell_k} v_{k,j} = 0$ for all $j = 1, \ldots, m_k$. Therefore, $(T - \lambda_k I)^{\ell_k}|_{V_k} = 0$, but $(T - \lambda_k I)^{\ell_k - 1} v_{k,\ell_k} = v_{k,1} \neq 0$. Since $V_k \subseteq G(\lambda_k, T)$, and $\dim V_k = \dim G(\lambda_k, T) = m_k$ by 8.31 and 8.23, we have $V_k = G(\lambda_k, T)$. By 5.27(b), the minimal polynomial of T is a product of the factors $(z - \lambda_k)$, and by Exercise 8B6, the exponents of these factors are exactly ℓ_k. Therefore, the minimal polynomial of T is p. □

Solution 8B.22 Let A_k and B_k be n_k-by-n_k matrices for each $k = 1, \ldots, m$. Let $N_k = n_1 + \cdots + n_{k-1}$ for each $k = 1, \ldots, m$, with the convention that $N_1 = 0$. Then the kth block matrix has indices i, j with $N_k + 1 \leq i, j \leq N_k + n_k$. Fix ℓ and s with $1 \leq s \leq n_\ell$, and consider the $(N_\ell + s)$th column of B. The $(N_\ell + s)$th column of B has zeros outside the ℓth block and equals the sth column of B_ℓ within the ℓth block. By 3.48 and 3.50, the $(N_\ell + s)$th column of AB is

$$\sum_{t=1}^{n_\ell} (B_\ell)_{t,s} A_{\cdot, N_\ell + t}.$$

Since A is block diagonal, every column $A_{\cdot, N_\ell + t}$ has zeros outside the rows $N_\ell + 1 \leq i \leq N_\ell + n_\ell$, and equals $(A_\ell)_{\cdot, t}$ within those rows. Then the $(N_\ell + s)$th column of AB is zero outside the ℓth block and equals

$$\sum_{t=1}^{n_\ell} (B_\ell)_{t,s} (A_\ell)_{\cdot, t} = A_\ell (B_\ell)_{\cdot, s} = (A_\ell B_\ell)_{\cdot, s}$$

by 3.48 and 3.50. Since this holds for any column of any block, we conclude that AB is block diagonal with blocks $A_k B_k$. □

Solution 8B.23

Axler Note. *See Exercise 3B33 for the definition of the complexification $T_{\mathbb{C}}$.*

Define $C : V_{\mathbb{C}} \to V_{\mathbb{C}}$ by $C(u + iv) = u - iv$ for $u, v \in V$. We claim that C is additive, conjugate linear, satisfies $C^2 = I$, and commutes with $T_{\mathbb{C}}$. To see additivity,

let $u_1 + iv_1, u_2 + iv_2 \in V_{\mathbb{C}}$. Then

$$\begin{aligned} C((u_1 + iv_1) + (u_2 + iv_2)) &= C((u_1 + u_2) + i(v_1 + v_2)) \\ &= (u_1 + u_2) - i(v_1 + v_2) \\ &= (u_1 - iv_1) + (u_2 - iv_2) \\ &= C(u_1 + iv_1) + C(u_2 + iv_2). \end{aligned}$$

To see conjugate linearity, let $\alpha = a + bi \in \mathbb{C}$ and $u + iv \in V_{\mathbb{C}}$. Then

$$\begin{aligned} C(\alpha(u + iv)) &= C((a + bi)(u + iv)) \\ &= C((au - bv) + i(av + bu)) \\ &= (au - bv) - i(av + bu) \\ &= (a - bi)(u - iv) \\ &= \overline{\alpha} C(u + iv). \end{aligned}$$

Moreover, $C^2(u + iv) = C(u - iv) = u + iv$, so $C^2 = I$. Finally,

$$CT_{\mathbb{C}}(u + iv) = C(Tu + iTv) = Tu - iTv = T_{\mathbb{C}}(u - iv) = T_{\mathbb{C}}C(u + iv),$$

so C commutes with $T_{\mathbb{C}}$.

(a) Since C commutes with $T_{\mathbb{C}}$ and $C(\lambda(u + iv)) = \overline{\lambda} C(u + iv)$ for any $u + iv \in V_{\mathbb{C}}$, we have $C(T_{\mathbb{C}} - \lambda I) = (T_{\mathbb{C}} - \overline{\lambda} I)C$. Then for any positive integer k, it follows by induction that $C(T_{\mathbb{C}} - \lambda I)^k = (T_{\mathbb{C}} - \overline{\lambda} I)^k C$. By definition, $u + iv \in G(\lambda, T_{\mathbb{C}})$ if and only if $(T_{\mathbb{C}} - \lambda I)^k (u + iv) = 0$ for some positive integer k. Moreover, $C^2 = I$ implies that C is injective, and additivity gives $C0 = 0$, so for any positive integer k, $(T_{\mathbb{C}} - \lambda I)^k (u + iv) = 0$ if and only if $(T_{\mathbb{C}} - \overline{\lambda} I)^k C(u + iv) = 0$. Since $C(u + iv) = u - iv$, we have $(T_{\mathbb{C}} - \overline{\lambda} I)^k (u - iv) = 0$ for some k if and only if $u - iv \in G(\overline{\lambda}, T_{\mathbb{C}})$.

(b) Suppose $v_1, \ldots, v_m$ is a basis of $G(\lambda, T_{\mathbb{C}})$. By part (a), $Cv_j \in G(\overline{\lambda}, T_{\mathbb{C}})$ for each $j = 1, \ldots, m$. We claim that $Cv_1, \ldots, Cv_m$ is a basis of $G(\overline{\lambda}, T_{\mathbb{C}})$. To see this, suppose there are scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{C}$ such that

$$\sum_{j=1}^{m} \alpha_j Cv_j = 0.$$

Applying C to both sides gives

$$C\left(\sum_{j=1}^{m} \alpha_j Cv_j\right) = \sum_{j=1}^{m} \overline{\alpha_j} C^2 v_j = \sum_{j=1}^{m} \overline{\alpha_j} v_j = C0 = 0.$$

By the linear independence of $v_1, \ldots, v_m$, we must have $\overline{\alpha_1} = \cdots = \overline{\alpha_m} = 0$, hence $\alpha_1 = \cdots = \alpha_m = 0$. Moreover, for any $w \in G(\overline{\lambda}, T_{\mathbb{C}})$, we have $Cw \in G(\lambda, T_{\mathbb{C}})$ by part (a), so $Cw = \sum_{j=1}^{m} \beta_j v_j$ for some $\beta_j \in \mathbb{C}$, hence

$$w = C(Cw) = C\left(\sum_{j=1}^{m} \beta_j v_j\right) = \sum_{j=1}^{m} \overline{\beta_j} Cv_j.$$

Therefore, $Cv_1, \ldots, Cv_m$ is a basis of $G(\overline{\lambda}, T_{\mathbb{C}})$, whence $\dim G(\overline{\lambda}, T_{\mathbb{C}}) = \dim G(\lambda, T_{\mathbb{C}})$. Thus, λ is an eigenvalue of $T_{\mathbb{C}}$ if and only if $\overline{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$.

(c) By Exercise 2B11, $\dim V_{\mathbb{C}} = \dim V$. Then $V_{\mathbb{C}}$ is nonzero with odd dimension, so we may apply 8.25 to $T_{\mathbb{C}}$ to conclude that $\dim V_{\mathbb{C}} = \sum_\mu d_\mu$, where μ ranges over the eigenvalues of $T_{\mathbb{C}}$ and d_μ is the multiplicity of μ. Note that for nonreal eigenvalues μ, we have $\mu \neq \overline{\mu}$. By part (b), $d_\mu = d_{\overline{\mu}}$ for each nonreal eigenvalue μ, so the sum of the multiplicities of nonreal eigenvalues is even. Since $\dim V_{\mathbb{C}}$ is odd, there must be at least one real eigenvalue of $T_{\mathbb{C}}$.

(d) By (c), $T_{\mathbb{C}}$ has a real eigenvalue $\nu \in \mathbb{R}$. By Exercise 5A17, ν is an eigenvalue of T. □

§8C

Solution 8C.1 Note that $T^2(z_1, z_2, z_3) = (z_3, 0, 0)$ and $T^3 = 0$. Suppose $S \in \mathrm{End}(\mathbb{C}^3)$ is such that $S^2 = T$. Then $S^6 = (S^2)^3 = T^3 = 0$, so S is nilpotent. By 8.16, $S^3 = 0$, but $T^2 = S^4 = SS^3 = 0$, which contradicts $T^2 \neq 0$. Therefore, T does not have a square root. □

Solution 8C.2

(a) With respect to the standard basis of $\mathbb{F}^5$, the matrix $\mathrm{Mat}(T)$ of T on and below the diagonal is 0, so T is nilpotent by 8.18.

(b) By part (a) and 8.16, $T^5 = 0$. As in the proof of 8.39, we may construct the polynomial

$$p(z) = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \frac{5}{128}z^4.$$

Squaring $p(z)$ gives

$$p^2(z) = 1 + z + z^5 q(z)$$

for some $q \in \mathbb{F}[X]$. Let $S = p(T)$, so $S^2 = I + T + T^5 q(T) = I + T$ by 5.17(a) and $T^5 = 0$.

□

Solution 8C.3

Manual Remark. *The coefficients in the recursion below are given by* $a_j = \binom{1/3}{j}$ *for* $j \geq 1$.

Let $T \in \mathrm{End}(V)$ be invertible with distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ and let $n = \dim V$. By 5.7, $\lambda_k \neq 0$ for each $k = 1, \ldots, m$. Let $G_k = G(\lambda_k, T)$ for each $k = 1, \ldots, m$. Then the generalized eigenspace decomposition of V is

$$V = \bigoplus_{k=1}^{m} G_k.$$

By 8.22(a), we may define $N_k = (T-\lambda_k I)|_{G_k} \in \mathrm{End}(G_k)$ for each $k = 1, \ldots, m$. Since N_k is nilpotent on G_k by 8.22(b), and $\dim G_k \leq n$, we have $N_k^n = 0$ by 8.16. As in the proof of 8.39 with a cubic in place of a square, set $a_1 = \frac{1}{3}$. For each $2 \leq j \leq n-1$, the coefficient of z^j in the expansion $(1 + a_1 z + \cdots + a_j z^j)^3$ is $3a_j + f_j(a_1, \ldots, a_{j-1})$ for some polynomial f_j in $a_1, \ldots, a_{j-1}$. Choosing a_j such that $3a_j + f_j(a_1, \ldots, a_{j-1}) = 0$, we can recursively determine $a_2, \ldots, a_{n-1}$. We then obtain the polynomial

$$p(z) = 1 + a_1 z + \cdots + a_{n-1} z^{n-1}$$

such that $p^3(z) = 1 + z + z^n r(z)$ for some $r \in \mathbb{F}[X]$. By the fundamental theorem of algebra, the polynomial $z^3 - \lambda_k$ has a zero in $\mathbb{C}$ for each $k = 1, \ldots, m$, say μ_k. Define the operators $S_k = \mu_k p(\lambda_k^{-1} N_k)$ on G_k for each k. By 5.17(a) and $N_k^n = 0$, it follows that $S_k^3 = (\mu_k p(\lambda_k^{-1} N_k))^3 = \lambda_k I + N_k = T|_{G_k}$. For each $v \in V$, by the generalized eigenspace decomposition, we may uniquely write $v = v_1 + \cdots + v_m$ with $v_k \in G_k$ for each $k = 1, \ldots, m$. Define $S \in \mathrm{End}(V)$ by $Sv = S_1 v_1 + \cdots + S_m v_m$. Since $Sv_k = S_k v_k \in G_k$ for each k, we have $S^3 v_k = S_k^3 v_k = T|_{G_k} v_k = T v_k$ for each k, and thus $S^3 = T$ on all of V. □

Solution 8C.4 $\Rightarrow$ Let $S \in \mathrm{End}(V)$ be such that $S^2 = -I$. Then $\ker(S^2 + 0S + I) = \ker(-I + I) = \ker(0) = V$. Since $0^2 < 4$, 5.33 gives that $\dim V$ is even.
$\Leftarrow$ Let $\dim V = 2n$ for some positive integer n, and let V have a basis $v_1, \ldots, v_{2n}$. By the linear map lemma, there exists a linear map $S \in \mathrm{End}(V)$ such that $Sv_{2k-1} = v_{2k}$ and $Sv_{2k} = -v_{2k-1}$ for each $k = 1, \ldots, n$. Then $S^2 v_{2k-1} = Sv_{2k} = -v_{2k-1}$ and $S^2 v_{2k} = S(-v_{2k-1}) = -v_{2k}$ for each $k = 1, \ldots, n$. Since S^2 and $-I$ agree on a basis of V, the uniqueness part of the linear map lemma implies that $S^2 = -I$. □

Solution 8C.5 Let $N = T - 2I$. Then $N(w, z) = (-3w - z, 9w + 3z) = (-3w - z)(1, -3)$, so

$$N^2(w, z) = (-3w - z)N(1, -3) = 0.$$

Let $v = (0, -1)$. Then $Nv = (1, -3) \neq 0$, so by Exercise 8A2, v, Nv is a linearly independent list of length 2, and thus is a basis of $\mathbb{C}^2$ by 2.38. We have $TNv = 2Nv$ and $Tv = Nv + 2v$, so

$$\mathrm{Mat}(T, (Nv, v)) = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

Therefore, the list $(1, -3), (0, -1)$ is a Jordan basis for T. □

Solution 8C.6 Define $v = \frac{x^4}{4!}$, so $D^4 v = 1$ and $D^5 v = 0$. By Exercise 8A2, the list $v, Dv, D^2 v, D^3 v, D^4 v$ is a linearly independent list of length 5 and thus forms a basis of $\mathbb{R}[X]_{\leq 4}$ by 2.38. Then

$$\mathrm{Mat}(D, (D^4 v, D^3 v, D^2 v, Dv, v)) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

so the list $1, x, \frac{x^2}{2!}, \frac{x^3}{3!}, \frac{x^4}{4!}$ is a Jordan basis for D. □

Solution 8C.7 Let $A_1, \dots, A_p$ be the blocks as in 8.44, where each A_k is d_k-by-d_k. Let $v_{k,1}, \dots, v_{k,d_k}$ be the basis vectors of the block A_k for each $k = 1, \dots, p$. By 8.44,

$$Tv_{k,j} = \begin{cases} \lambda_k v_{k,1} & \text{if } j = 1, \\ \lambda_k v_{k,j} + v_{k,j-1} & \text{if } j > 1. \end{cases}$$

Note that $Tv_{k,1} = \lambda_k v_{k,1}$ with $v_{k,1} \neq 0$. Since T is nilpotent, 8.17(a) implies that $\lambda_k = 0$ for all k. Then $Tv_{k,1} = 0$ and $Tv_{k,j} = v_{k,j-1}$ for $j > 1$, so

$$T^{j-1}v_{k,j} = v_{k,1}, \quad T^j v_{k,j} = 0.$$

The entries above the diagonal between the blocks A_k and A_{k+1} are all 0, so the consecutive string of 1's cannot extend beyond a single block. Therefore, each block A_k contributes a consecutive string of 1's of length $d_k - 1$, so $m + 1 = \max_k d_k$. By 8.18, the minimal polynomial of T is z^r for some r. We now claim that $r = m + 1$.

$\leq$ Since each $v_{k,j}$ has $j \leq d_k \leq m + 1$, we have $T^{m+1}v_{k,j} = T^{m+1-j}(T^j v_{k,j}) = 0$. Therefore, $T^{m+1} = 0$, and 5.29 yields $r \leq m + 1$.

$\geq$ Consider some k such that $d_k = m + 1$. Then $T^m v_{k,m+1} = v_{k,1} \neq 0$, so $T^m \neq 0$. Therefore, $r \geq m + 1$ by 5.29. Combining this with the previous inequality gives $r = m + 1$. □

Solution 8C.8 Let $A_1, \dots, A_p$ be the blocks of $\mathrm{Mat}(T)$ with respect to the basis $v_1, \dots, v_n$ and let λ_k be the eigenvalue corresponding to the block A_k. By 3.43 and Exercise 8B22, $\mathrm{Mat}(T^2) = \mathrm{Mat}(T)^2$ is block diagonal with blocks $A_1^2, \dots, A_p^2$. Write $A_k = \lambda_k I + N_k$, where N_k is the matrix with 1's on the superdiagonal and 0's elsewhere. Then $A_k^2 = \lambda_k^2 I + 2\lambda_k N_k + N_k^2$, and N_k^2 has 1's on the second superdiagonal and 0's elsewhere by 3.46. Therefore, each block A_k^2 has λ_k^2 on the diagonal, $2\lambda_k$ on the superdiagonal, 1's on the second superdiagonal, and 0's elsewhere. □

Solution 8C.9 By 8.45, nilpotence of T gives a Jordan basis $u_1, \dots, u_{\dim V}$ for T. Let $A_1, \dots, A_n$ be the blocks of $\mathrm{Mat}(T, (u_1, \dots, u_{\dim V}))$ with diagonal entries $\lambda_1, \dots, \lambda_n$. Let $u_{k,1}, \dots, u_{k,d_k}$ be the basis vectors corresponding to the d_k-by-d_k block A_k. By 5.41 and 8.17(a), $\lambda_k = 0$, so $Tu_{k,1} = 0$ and $Tu_{k,j} = u_{k,j-1}$ for $2 \leq j \leq d_k$ by 8.44. Let $v_k = u_{k,d_k}$ and $m_k = d_k - 1$ for $1 \leq k \leq n$.
(a) is satisfied because $T^j v_k = u_{k,d_k-j}$ for $0 \leq j \leq m_k$, so the vectors $T^{m_k}v_k, \dots, Tv_k, v_k$ are exactly $u_{k,1}, \dots, u_{k,d_k}$ in order.

(b) is satisfied because $T^{m_k+1}v_k = T^{d_k}u_{k,d_k} = 0$ for each k, as $u_{k,1} \in \ker T$. □

Solution 8C.10 Let $A_1, \dots, A_p$ be the Jordan blocks of $\mathrm{Mat}(T, (v_1, \dots, v_n))$ with diagonal entries $\lambda_1, \dots, \lambda_p$. Let $v_{k,1}, \dots, v_{k,d_k}$ be the basis vectors among $v_1, \dots, v_n$ corresponding to A_k in order. By 8.44, $Tv_{k,1} = \lambda_k v_{k,1}$ and $Tv_{k,j} = v_{k,j-1} + \lambda_k v_{k,j}$ for $2 \leq j \leq d_k$. In the basis $v_n, \dots, v_1$, the basis vectors corresponding to A_k appear in the reverse order $v_{k,d_k}, \dots, v_{k,1}$, with the blocks appearing in the order $A_p, \dots, A_1$. Writing $w_\ell = v_{k,d_k+1-\ell}$ for $1 \leq \ell \leq d_k$, we have $Tw_\ell = \lambda_k w_\ell + w_{\ell+1}$ for $1 \leq \ell \leq d_k - 1$ and $Tw_{d_k} = \lambda_k w_{d_k}$. Therefore, $\mathrm{Mat}(T, (v_n, \dots, v_1))$ is block diagonal with the blocks $A_p^{\mathrm{t}}, \dots, A_1^{\mathrm{t}}$. Each block A_k^{t} has λ_k on its diagonal, 1's on its subdiagonal, and 0's elsewhere. □

Solution 8C.11 Let $v_1, \ldots, v_n$ be a Jordan basis for T. Let $A_1, \ldots, A_p$ be the blocks of $\mathrm{Mat}(T, (v_1, \ldots, v_n))$ with diagonal entries $\lambda_1, \ldots, \lambda_p$. Let $v_{k,1}, \ldots, v_{k,d_k}$ be the basis vectors corresponding to A_k in order. By 8.44, $(T - \lambda_k I)v_{k,1} = 0$ and $(T - \lambda_k I)v_{k,j} = v_{k,j-1}$ for $2 \leq j \leq d_k$. Hence, λ_k is an eigenvalue of T and $(T - \lambda_k I)^j v_{k,j} = 0$ for $1 \leq j \leq d_k$. Since $v_{k,j} \neq 0$, it follows that $v_{k,j}$ is a generalized eigenvector of T corresponding to λ_k. □

Solution 8C.12 Let $v_1, \ldots, v_n$ be a Jordan basis for T. Let $A_1, \ldots, A_p$ be the blocks of $\mathrm{Mat}(T, (v_1, \ldots, v_n))$ with diagonal entries $\lambda_1, \ldots, \lambda_p$. Let $v_{k,1}, \ldots, v_{k,d_k}$ be the basis vectors corresponding to the block A_k, and define $U_k = \mathrm{span}(v_{k,1}, \ldots, v_{k,d_k})$. By 8.44, T maps the vectors $v_{k,1}, \ldots, v_{k,d_k}$ into U_k, so U_k is invariant under T, and $\mathrm{Mat}(T|_{U_k}, (v_{k,1}, \ldots, v_{k,d_k})) = A_k$. By 5.65, $T|_{U_k}$ is diagonalizable for each k. Since A_k is upper triangular with λ_k on the diagonal, 5.41 yields that λ_k is the only eigenvalue of $T|_{U_k}$. By 5.55, $U_k = E(\lambda_k, T|_{U_k})$, so $T|_{U_k} = \lambda_k I$, and hence $A_k = \lambda_k I$. Since this holds for each block A_k, it follows that $\mathrm{Mat}(T, (v_1, \ldots, v_n))$ is diagonal. □

Solution 8C.13

Axler Note. *This exercise shows that $n = \dim \ker T$. Thus, the positive integer n that appears above depends only on T and not on the specific Jordan basis chosen for T.*

Since $T^{m_k+1} v_k = 0$ for each k, $T^{m_k} v_k \in \ker T$ for each k. Observe that $T(T^j v_k) = T^{j+1} v_k$ for each k and $0 \leq j \leq m_k$, where $T^{m_k+1} v_k = 0$. Therefore, the list $Tv_1, \ldots, T^{m_1} v_1, \ldots, Tv_n, \ldots, T^{m_n} v_n$ spans $\mathrm{im}\, T$ and is linearly independent since it is part of the given basis. Consequently, the list forms a basis of $\mathrm{im}\, T$, and hence $\dim \mathrm{im}\, T = \sum_{k=1}^{n} m_k$. Because $\dim V = \sum_{k=1}^{n} (m_k + 1)$, the fundamental theorem of linear maps gives

$$\dim \ker T = \sum_{k=1}^{n} (m_k + 1) - \sum_{k=1}^{n} m_k = n.$$

Since $T^{m_1} v_1, \ldots, T^{m_n} v_n$ is part of the given basis, it is a linearly independent list of length n in $\ker T$, wherefore it forms a basis of $\ker T$ by 2.38. □

Solution 8C.14 Let p be the minimal polynomial of T and let $n = \dim V$.

$\Rightarrow$ Suppose no direct sum decomposition of V into two nonzero subspaces invariant under T exists. By 8.46, there exists a Jordan basis $v_1, \ldots, v_n$ of V with blocks $A_1, \ldots, A_q$. Define U to be the span of the basis vectors corresponding to the first block A_1 and W to be the span of the remaining basis vectors. By 8.44, T maps the basis vectors corresponding to each block into their own span, so U and W are invariant under T. Since $U + W$ contains the basis vectors and $\dim U + \dim W = n$, it follows that $V = U \oplus W$ by 3.94. Since $U \neq \{0\}$, the hypothesis implies that $W = \{0\}$. Hence, $\mathrm{Mat}(T, (v_1, \ldots, v_n)) = A_1$ is a single n-by-n block. Let $\lambda = \lambda_1$. By 8.44, $(T - \lambda I)v_1 = 0$ and $(T - \lambda I)v_j = v_{j-1}$ for $2 \leq j \leq n$, wherefore $(T - \lambda I)^n = 0$ and $(T - \lambda I)^{n-1} v_n = v_1 \neq 0$. By 5.29, p divides $(z - \lambda)^n$, so $p = (z - \lambda)^r$ for some positive integer $r \leq n$. On the other hand, if $r \leq n - 1$, then $(z - \lambda)^{n-1}$ is a multiple of p, so $(T - \lambda I)^{n-1} = 0$ by 5.29, contradicting $(T - \lambda I)^{n-1} v_n = v_1 \neq 0$. Hence, $r = n$, so $p = (z - \lambda)^n$.

$\Leftarrow$ Suppose $p = (z-\lambda)^n$ for some $\lambda \in \mathbb{C}$. Let $V = U \oplus W$ with U and W invariant under T. Since $(T-\lambda I)^n = 0$ and invariance of U under T gives invariance under $T-\lambda I$, it follows that $(T-\lambda I)^{\dim U} = 0$ on U by 8.16. A similar argument shows that $(T-\lambda I)^{\dim W} = 0$ on W. Let $M = \max(\dim U, \dim W)$. Then $(T-\lambda I)^M = 0$ on $U+W = V$, so $M \geq n$ by 5.29. By 3.94, $\dim U + \dim W = n$, so $M \geq \dim U + \dim W$, forcing $\min(\dim U, \dim W) = 0$. Hence, either $U = \{0\}$ or $W = \{0\}$, proving that no direct sum decomposition into two nonzero invariant subspaces exists. □

§8D

Solution 8D.1 Let $n = \dim V$. By 6.35, there exists an orthonormal basis $e_1, \ldots, e_n$ of V. By 8.55, homogeneity in the first slot, conjugate symmetry of the inner product, and 6.30(c), we have

$$\operatorname{tr} T = \sum_{k=1}^{n} \langle Te_k, e_k \rangle = \sum_{k=1}^{n} \langle \langle e_k, v \rangle w, e_k \rangle = \sum_{k=1}^{n} \langle w, e_k \rangle \overline{\langle v, e_k \rangle} = \langle w, v \rangle. \qquad \square$$

Solution 8D.2 Let $n = \dim V$. By Exercise 3B27, $V = \ker P \oplus \operatorname{im} P$. Let $v_1, \ldots, v_m$ be a basis of $\operatorname{im} P$ and $u_1, \ldots, u_{n-m}$ be a basis of $\ker P$, where $n-m = \dim \ker P$ from the fundamental theorem of linear maps. Since the list $v_1, \ldots, v_m, u_1, \ldots, u_{n-m}$ spans $\ker P + \operatorname{im} P = V$ and is of length n, it is a basis of V by 2.42. For each $v_k \in \operatorname{im} P$, there is $w \in V$ such that $v_k = Pw$, and hence $Pv_k = P^2 w = Pw = v_k$. Since $Pu_j = 0$ for $j = 1, \ldots, n-m$, the matrix of P with respect to this basis is diagonal with m 1's and $n-m$ 0's on the diagonal. Therefore, $\operatorname{tr} P = m = \dim \operatorname{im} P$. □

Solution 8D.3 Let $n = \dim V$. We show the existence of an operator $S \in \operatorname{End}(\mathbb{C}^n)$ such that $S^5 = S$ and $\operatorname{tr} S = \operatorname{tr} T$. Indeed, fix a basis $v_1, \ldots, v_n$ of V, and let $A = \operatorname{Mat}(T, (v_1, \ldots, v_n))$. Viewing $A \in \mathbb{C}^{n,n}$, the linear map lemma yields a unique $S \in \operatorname{End}(\mathbb{C}^n)$ such that $\operatorname{Mat}(S)$ with respect to the standard basis of $\mathbb{C}^n$ is A, so $\operatorname{tr} T = \operatorname{tr} A = \operatorname{tr} S$. 3.43 yields

$$\operatorname{Mat}(S^5) = A^5 = \operatorname{Mat}(T^5) = \operatorname{Mat}(T) = A = \operatorname{Mat}(S).$$

By 3.71, $S^5 = S$. Suppose $Sv = \lambda v$ for $v \neq 0$ and $\lambda \in \mathbb{C}$. Then $S^5 v = \lambda^5 v = Sv = \lambda v$, so $\lambda^5 = \lambda$. Because $z^5 - z = z(z-1)(z+1)(z-i)(z+i)$ has zeros $0, 1, -1, i, -i$, we see that the eigenvalues of S are all among these numbers. Letting d_λ be the multiplicity of the eigenvalue λ, where $d_\lambda = 0$ if λ is not an eigenvalue, we have by 8.52 that

$$\operatorname{tr} S = (d_1 - d_{-1}) + (d_i - d_{-i})i.$$

Therefore, both $\operatorname{Re}(\operatorname{tr} S) = d_1 - d_{-1}$ and $\operatorname{Im}(\operatorname{tr} S) = d_i - d_{-i}$ are integers. Since $\operatorname{tr} T = \operatorname{tr} S$, the real and imaginary parts of $\operatorname{tr} T$ are also integers. □

Solution 8D.4 Let $n = \dim V$. By 6.35, there exists an orthonormal basis $e_1, \dots, e_n$ of V. By 8.55, we have

$$\operatorname{tr} T^* = \sum_{k=1}^{n} \langle T^* e_k, e_k \rangle = \sum_{k=1}^{n} \langle e_k, T e_k \rangle = \sum_{k=1}^{n} \overline{\langle T e_k, e_k \rangle} = \overline{\sum_{k=1}^{n} \langle T e_k, e_k \rangle} = \overline{\operatorname{tr} T}. \quad \square$$

Solution 8D.5 Let $n = \dim V$. By 7.38(c), there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $\operatorname{Mat}(T, (e_1, \dots, e_n))$ is diagonal with nonnegative diagonal entries $\lambda_1, \dots, \lambda_n$. Since $0 = \operatorname{tr} T = \sum_{k=1}^{n} \lambda_k$ and each $\lambda_k \geq 0$, we must have $\lambda_k = 0$ for all k. Therefore, $\operatorname{Mat}(T, (e_1, \dots, e_n)) = 0$, so $T = 0$. $\square$

Solution 8D.6 Let $n = \dim V$. By 6.57(g), $P^2 = P$ and $Q^2 = Q$, so P and Q are self-adjoint by Exercise 7A20. Since $Q^2 = Q$, we have $\operatorname{tr}(PQ) = \operatorname{tr}((PQ)Q)$, so 8.56 gives $\operatorname{tr}((PQ)Q) = \operatorname{tr}(Q(PQ)) = \operatorname{tr}(QPQ)$. By 7.5(d), $QPQ = QPPQ = (PQ)^*(PQ)$. 6.35 yields an orthonormal basis $e_1, \dots, e_n$ of V, so 8.55 gives

$$\operatorname{tr}(PQ) = \sum_{k=1}^{n} \langle (PQ)^*(PQ) e_k, e_k \rangle = \sum_{k=1}^{n} \| PQ e_k \|^2 \geq 0. \quad \square$$

Solution 8D.7 Note that $\operatorname{tr} T = 51 + (-40) + 1 = 12$. By 8.25, the eigenvalues of T, repeated with multiplicity, form a list of length 3 that contains $-48, 24$, and a third eigenvalue λ. 8.52 gives $-48 + 24 + \lambda = 12$, so $\lambda = 36$. $\square$

Solution 8D.8 This is false. Take $V = \mathbb{F}^2$ with $S = T = I$, so $ST = I$. Then $\operatorname{tr}(ST) = 2$, but $(\operatorname{tr} S)(\operatorname{tr} T) = 4$. $\square$

Solution 8D.9 Let $v_1, \dots, v_n$ be a basis of V, and let $A = \operatorname{Mat}(T, (v_1, \dots, v_n))$. By the linear map lemma, there exist operators $S_{j,k} \in \operatorname{End}(V)$ such that $E_{j,k} = \operatorname{Mat}(S_{j,k}, (v_1, \dots, v_n))$ has a 1 in the (j,k)th entry and 0 elsewhere. By 3.46,

$$(E_{j,k} A)_{\ell,\ell} = \sum_{m=1}^{n} (E_{j,k})_{\ell,m} A_{m,\ell} = (E_{j,k})_{\ell,k} A_{k,\ell} = \begin{cases} A_{k,j} & \text{if } \ell = j, \\ 0 & \text{otherwise.} \end{cases}$$

By 3.43, $\operatorname{tr}(S_{j,k} T) = \operatorname{tr}(E_{j,k} A) = A_{k,j}$. By assumption, $\operatorname{tr}(S_{j,k} T) = 0$ for all j, k, so $A_{k,j} = 0$ for all j, k. Therefore, $A = 0$, so by the linear map lemma, $T = 0$. $\square$

Solution 8D.10

Axler Hint. *Suppose that $v_1, \dots, v_n$ is a basis of V. For $j, k \in \{1, \dots, n\}$, define $P_{j,k} \in \operatorname{End}(V)$ by $P_{j,k}(a_1 v_1 + \cdots + a_n v_n) = a_k v_j$. Prove that*

$$\tau(P_{j,k}) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Then for $T \in \operatorname{End}(V)$, use the equation $T = \sum_{k=1}^{n} \sum_{j=1}^{n} \operatorname{Mat}(T)_{j,k} P_{j,k}$ to show that $\tau(T) = \operatorname{tr} T$.

To show existence of such a linear functional, observe that 8.56 gives linearity of the trace and that $\operatorname{tr}(ST) = \operatorname{tr}(TS)$ for all $S, T \in \operatorname{End}(V)$. Since $\operatorname{Mat}(I) = I$, we have $\operatorname{tr} I = \dim V$, so the trace satisfies the same conditions as τ.

To show uniqueness, let $n = \dim V$ and let $v_1, \ldots, v_n$ be a basis of V. Let τ be a linear functional on $\operatorname{End}(V)$ such that $\tau(ST) = \tau(TS)$ for all $S, T \in \operatorname{End}(V)$ and $\tau(I) = n$. Let $P_{j,k}$ be as in the hint. Then for $a_1, \ldots, a_n \in \mathbb{F}$, we have

$$P_{j,k} P_{k,j}(a_1 v_1 + \cdots + a_n v_n) = P_{j,k}(a_j v_k) = a_j v_j = P_{j,j}(a_1 v_1 + \cdots + a_n v_n).$$

Therefore, $P_{j,k} P_{k,j} = P_{j,j}$. Swapping j and k gives $P_{k,j} P_{j,k} = P_{k,k}$. Since $\tau(ST) = \tau(TS)$ for all $S, T \in \operatorname{End}(V)$, we have

$$\tau(P_{j,j}) = \tau(P_{j,k} P_{k,j}) = \tau(P_{k,j} P_{j,k}) = \tau(P_{k,k})$$

for all j, k, so $\tau(P_{j,j})$ is the same for all j. Moreover,

$$I = \sum_{j=1}^{n} P_{j,j},$$

so $n = \tau(I) = n\tau(P_{1,1})$, whence $\tau(P_{1,1}) = 1$. Thus, $\tau(P_{j,j}) = 1$ for all j.

Fix $T \in \operatorname{End}(V)$, and let $A = \operatorname{Mat}(T)$ be the matrix of T with respect to the basis $v_1, \ldots, v_n$. Then for any $v = a_1 v_1 + \cdots + a_n v_n \in V$, we have

$$P_{j,j} T P_{j,j} v = a_j P_{j,j} \left(\sum_{\ell=1}^{n} A_{\ell,j} v_\ell \right) = a_j A_{j,j} v_j = A_{j,j} P_{j,j} v,$$

so $P_{j,j} T P_{j,j} = A_{j,j} P_{j,j}$. From $I = \sum_{k=1}^{n} P_{k,k}$, we get $T = \sum_{k=1}^{n} T P_{k,k}$. Then

$$\tau(T) = \sum_{k=1}^{n} \tau((T P_{k,k}) P_{k,k}) = \sum_{k=1}^{n} \tau(P_{k,k} T P_{k,k}) = \sum_{k=1}^{n} A_{k,k} \tau(P_{k,k}) = \sum_{k=1}^{n} A_{k,k} = \operatorname{tr} T,$$

where the first equality follows from linearity and $P_{k,k}^2 = P_{k,k}$; the second equality follows from the property $\tau(ST) = \tau(TS)$; the third equality follows from $P_{k,k} T P_{k,k} = A_{k,k} P_{k,k}$ and linearity; the fourth equality follows from $\tau(P_{k,k}) = 1$; and the fifth equality is by the definition of the trace. □

Solution 8D.11

Axler Note. *The numbers $\langle Te_k, f_j \rangle$ are the entries of the matrix of T with respect to the orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$. These numbers depend on the bases, but $\operatorname{tr}(T^*T)$ does not depend on a choice of bases. Thus, this exercise shows that the sum of the squares of the absolute values of the matrix entries does not depend on which orthonormal bases are used.*

Apply 8.55 to $T^*T \in \operatorname{End}(V)$. Then the definition of the adjoint and 6.30(b) in W yield

$$\operatorname{tr}(T^*T) = \sum_{k=1}^{n} \langle T^*Te_k, e_k \rangle = \sum_{k=1}^{n} \|Te_k\|^2 = \sum_{k=1}^{n} \sum_{j=1}^{m} |\langle Te_k, f_j \rangle|^2.$$

□

Solution 8D.12

Axler Note. ***Caution:*** *The norm of a linear map $T \in \operatorname{Hom}(V, W)$ as defined by 7.86 is not the same as the norm that comes from the inner product in (a) above. Unless explicitly stated otherwise, always assume that $\|T\|$ refers to the norm as defined by 7.86. The norm that comes from the inner product in (a) is called the* Frobenius norm *or the* Hilbert–Schmidt norm.

Let $n = \dim V$ and $m = \dim W$. By 6.35, there exist orthonormal bases $e_1, \dots, e_n$ of V and $f_1, \dots, f_m$ of W. Define the map $\phi\colon \operatorname{Hom}(V, W) \to \mathbb{F}^{mn}$ by mapping T to $\operatorname{Mat}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$ as a vector in $\mathbb{F}^{mn}$, where the entries are listed column by column. 3.71 gives that ϕ is linear and injective. By 3.43 and 7.9, $\operatorname{Mat}(T^*S) = \operatorname{Mat}(T)^* \operatorname{Mat}(S)$ with respect to the fixed orthonormal bases, so

$$\operatorname{tr}(T^*S) = \sum_{k=1}^{n} \sum_{j=1}^{m} \operatorname{Mat}(S)_{j,k} \overline{\operatorname{Mat}(T)_{j,k}} = \langle \phi(S), \phi(T) \rangle,$$

where the right-hand side is the Euclidean inner product on $\mathbb{F}^{mn}$.

(a) Note that positivity, additivity, homogeneity in the first slot, and conjugate symmetry of the inner product on $\operatorname{Hom}(V, W)$ follow from the corresponding properties of the Euclidean inner product on $\mathbb{F}^{mn}$ via the map ϕ. To see definiteness, note that $\langle 0, 0 \rangle = 0$ and if $\langle T, T \rangle = 0$, then $\langle \phi(T), \phi(T) \rangle = 0$, which implies $\phi(T) = 0$, and hence $T = 0$ by injectivity of ϕ.

(b) This is exactly the identity computed in the display above, which shows that the inner product on $\operatorname{Hom}(V, W)$ corresponds to the Euclidean inner product on $\mathbb{F}^{mn}$ under the map ϕ. □

Solution 8D.13 Let $S, T \in \operatorname{End}(\mathbb{F}[X])$ be defined by $Sf = f'$ and $Tf = zf$ for all $f \in \mathbb{F}[X]$. Each f is a linear combination of monomials z^k, so it suffices to check the identity on $f = z^k$ for each $k \geq 0$. We have $STz^k = S(z^{k+1}) = (k+1)z^k$ for $k \geq 0$. Also, $TSz^k = T(kz^{k-1}) = kz^k$ holds for $k \geq 1$, and for $k = 0$, $TSz^0 = T(Sz^0) = T(0) = 0$. Therefore, for all $k \geq 0$, we have $(ST - TS)z^k = z^k$, which shows that $ST - TS = I$ on $\mathbb{F}[X]$. □

Chapter 9

§9A

Solution 9A.1 Let $c = \beta(1, 1)$. Then for all $x, y \in \mathbb{F}$, homogeneity in the first and second slots gives

$$\beta(x, y) = \beta(x \cdot 1, y \cdot 1) = xy\beta(1, 1) = cxy.$$ □

Solution 9A.2

Axler Note. *This exercise shows that if $n = \dim V$, then every bilinear form on V is of the form given by the last bullet point of Example 9.2.*

Let $v_1, \ldots, v_n$ be a basis of V with dual basis $\phi_1, \ldots, \phi_n$ of V'. For any $u \in V$, 3.114 yields $u = \phi_1(u)v_1 + \cdots + \phi_n(u)v_n$. For each $k = 1, \ldots, n$, define $\tau_k \in V'$ by $\tau_k(v) = \beta(v_k, v)$. Then for any $v \in V$,

$$\beta(u, v) = \sum_{k=1}^{n} \phi_k(u)\beta(v_k, v) = \sum_{k=1}^{n} \phi_k(u)\tau_k(v). \qquad \square$$

Solution 9A.3 For any $u, v \in V$, we have $(u, v) = (u, 0) + (0, v)$, so linearity of β on $V \times V$ gives $\beta(u, v) = \beta(u, 0) + \beta(0, v)$. Applying 3.10 to $w \mapsto \beta(u, w)$ and $w \mapsto \beta(w, v)$ gives $\beta(u, 0) = 0$ and $\beta(0, v) = 0$, respectively. Hence, $\beta(u, v) = \beta(u, 0) + \beta(0, v) = 0$ for all $u, v \in V$, so $\beta = 0$. $\square$

Solution 9A.4

Axler Note. *This exercise states that if V is a real inner product space, then every bilinear form on V is of the form given by the third bullet point in 9.2.*

Appendix. *An alternate proof is given at the end of this solution.*

For each $v \in V$, define $\phi_v \in V'$ by $\phi_v(u) = \langle u, v \rangle$. By Exercise 6C13, $\Phi: V \to V'$ defined by $\Phi(v) = \phi_v$ is an isomorphism, so $\Phi^{-1} \in \operatorname{Hom}(V', V)$ exists. Define $\Psi: V \to V'$ by $\Psi(v)(u) = \beta(u, v)$ for all $u \in V$. By linearity of β in the first slot, $\Psi(v) \in V'$. Moreover, for any $\lambda, \mu \in \mathbb{R}$ and $v, w \in V$, we have

$$\Psi(\lambda v + \mu w)(u) = \beta(u, \lambda v + \mu w) = \lambda\Psi(v)(u) + \mu\Psi(w)(u)$$

for all $u \in V$, so $\Psi \in \operatorname{Hom}(V, V')$. For any $S \in \operatorname{End}(V)$, we have

$$\begin{aligned} \beta(u, v) = \langle u, Sv \rangle \text{ for all } u, v \in V &\iff \Psi(v) = \phi_{Sv} = \Phi(Sv) \text{ for all } v \in V \\ &\iff \Psi = \Phi S \\ &\iff S = \Phi^{-1}\Psi. \end{aligned}$$

Thus, the operator $T = \Phi^{-1}\Psi$ satisfies $\beta(u, v) = \langle u, Tv \rangle$ for all $u, v \in V$, and it is unique.

Alternate proof. This proof uses the Riesz representation theorem to each functional $u \mapsto \beta(u, v)$ separately and describes, in detail, where the assumption $\mathbb{F} = \mathbb{R}$ is used.

Alternate proof. Civil Service Pigeon. For each $v \in V$, the map $u \mapsto \beta(u, v)$ is a linear functional on V. By the Riesz representation theorem, there exists a unique vector $Tv \in V$ such that $\beta(u, v) = \langle u, Tv \rangle$ for all $u \in V$. This defines a map $T : V \to V$.

For any $v, w \in V$ and $\lambda, \mu \in \mathbb{R}$ and $u \in V$, we have

$$\langle u, T(\lambda v + \mu w) \rangle = \beta(u, \lambda v + \mu w) = \lambda\beta(u, v) + \mu\beta(u, w) = \langle u, \lambda Tv + \mu Tw \rangle .$$

Since this holds for all $u \in V$, it follows that $T(\lambda v + \mu w) = \lambda Tv + \mu Tw$, so $T \in \mathcal{L}(V)$.

To prove uniqueness, if $S \in \mathcal{L}(V)$ also satisfies $\beta(u, v) = \langle u, Sv\rangle$ for all $u, v \in V$, then $\langle u, Tv - Sv\rangle = 0$ for all $u \in V$, which implies $Tv = Sv$ for all $v \in V$ by choosing $u = Tv - Sv$. Hence, $T = S$. □

Solution 9A.5 Let $\langle \cdot, \cdot \rangle_T$ be defined as in Exercise 7C23. Since $\mathbb{F} = \mathbb{R}$, for all $u, v \in V$, we have

$$\langle u, v\rangle_T = \langle Tu, v\rangle = \langle v, Tu\rangle = \beta(v, u),$$

so $\langle \cdot, \cdot \rangle_T$ is β with swapped arguments. Then β is an inner product if and only if $\langle \cdot, \cdot \rangle_T$ is an inner product, which holds if and only if T is an invertible positive operator, where the second equivalence follows from Exercise 7C23(a). □

Solution 9A.6 The statement is false. Consider $V = \mathbb{F}^2$ and let e_1, e_2 be the standard basis of V. Define $\rho: V \times V \to \mathbb{F}$ by

$$\rho(ae_1 + be_2, ce_1 + de_2) = ad + bc.$$

Then ρ is symmetric and bilinear. However, $\rho(e_1, e_1) = \rho(e_2, e_2) = 0$, but $\rho(e_1 + e_2, e_1 + e_2) = 2 \neq 0$. Therefore, the set $\{v \in V : \rho(v, v) = 0\}$ is not a subspace. □

Solution 9A.7

Manual Remark. *I ultimately decided to point out one specific failure within the proof, but there are other places where the argument relies on the entries being real.*

The proof of 9.13 fails when it claims that T is self-adjoint: since $f_1, \ldots, f_n$ is an orthonormal basis of V, 7.9 gives $\mathrm{Mat}(T^*) = B^* = \overline{B}$ because $B^{\mathrm{t}} = B$. Therefore, $T = T^*$ only holds when $B = \overline{B}$, i.e., when the entries of B are real, which fails when B is symmetric with non-real entries. □

Solution 9A.8 Let $n = \dim V$ and fix some basis of V. Let $\phi: V^{(2)} \to \mathbb{F}^{n,n}$ be defined by $\phi(\beta) = \mathrm{Mat}(\beta)$. Define $\mathcal{S}$ to be the set of symmetric n-by-n matrices over $\mathbb{F}$. By 9.5 and 9.12, ϕ restricts to an isomorphism from $V^{(2)}_{\mathrm{sym}}$ to $\mathcal{S}$, so $\dim V^{(2)}_{\mathrm{sym}} = \dim \mathcal{S}$. Consider the matrix $E_{j,k}$ with a 1 in the (j, k)th entry and 0 elsewhere. We claim that the set

$$\mathcal{C} = \{E_{j,j} : 1 \leq j \leq n\} \cup \{E_{j,k} + E_{k,j} : 1 \leq j < k \leq n\}$$

is a basis of $\mathcal{S}$. Indeed, each $A \in \mathcal{S}$ can be written as

$$A = \sum_{j=1}^{n} A_{j,j} E_{j,j} + \sum_{1 \leq j < k \leq n} A_{j,k}(E_{j,k} + E_{k,j}),$$

and comparing the (j, k)th entries, we see that the coefficients are uniquely determined, so 2.28 gives that $\mathcal{C}$ is a basis of $\mathcal{S}$. Counting the elements of $\mathcal{C}$, we have n matrices of the form $E_{j,j}$ and $\frac{n(n-1)}{2}$ matrices of the form $E_{j,k} + E_{k,j}$ with $j < k$. Hence,

$$\dim \mathcal{S} = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2} = \dim V^{(2)}_{\mathrm{sym}}.$$

Finally, 9.5 gives that $\dim V^{(2)} = n^2$ and 9.17 gives that $V^{(2)} = V^{(2)}_{\text{sym}} \oplus V^{(2)}_{\text{alt}}$, so by 3.94,

$$\dim V^{(2)}_{\text{alt}} = \dim V^{(2)} - \dim V^{(2)}_{\text{sym}} = n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}. \qquad \square$$

Solution 9A.9 For fixed $q \in V$, the map $p \mapsto \alpha(p,q)$ is linear because the integral is linear. Similarly, for fixed $p \in V$, the map $q \mapsto \alpha(p,q)$ is linear because differentiation and the integral are linear, and thus α is bilinear. Finally, for any $r \in V$, we have

$$\alpha(r,r) = \int_0^1 rr' = \frac{1}{2}\int_0^1 (r^2)' = \frac{1}{2}(r(1)^2 - r(0)^2) = 0,$$

since $r(0) = r(1)$. Hence, α is alternating. $\square$

Solution 9A.10 Let $U = \{p \in \mathbb{R}[X]_{\le n} : p(0) = p(1)\}$ and let α be as in Exercise 9A9. For any $p, q \in V$, we have $p, q' \in U$ because $q' \in \mathbb{R}[X]_{\le n-1}$ and $q'(0) = q'(1)$. Therefore, $\rho(p,q) = \alpha(p,q')$. Moreover, ρ is linear in p because α is linear in its first slot. Also, ρ is linear in q since $q \mapsto q'$ is linear and α is linear in its second slot. Finally, for any $p, q \in V$, 9.16 gives

$$\rho(p,q) = \alpha(p,q') = -\alpha(q',p) = -\int_0^1 q'p',$$

and swapping p and q results in the same expression, so ρ is symmetric. $\square$

§9B

Solution 9B.1 Let $n = \dim V$. Let $v_1, \dots, v_n$ be a basis of V and let $\phi_1, \dots, \phi_n$ be the dual basis. For each ordered m-tuple $(k_1, \dots, k_m) \in \{1, \dots, n\}^m$, define

$$\beta_{k_1,\dots,k_m}(u_1, \dots, u_m) = \phi_{k_1}(u_1) \cdots \phi_{k_m}(u_m).$$

The ℓth map $u_\ell \mapsto \beta_{k_1,\dots,k_m}(u_1, \dots, u_m)$ is linear because it is the product of $\phi_{k_\ell}(u_\ell)$, which is linear in u_ℓ, and the other factors, which are independent of u_ℓ. Therefore, $\beta_{k_1,\dots,k_m}$ is m-linear. We claim that the list $\mathcal{B} = (\beta_{k_1,\dots,k_m})_{(k_1,\dots,k_m)\in\{1,\dots,n\}^m}$ forms a basis of $V^{(m)}$.

Indeed, suppose there are scalars $c_{k_1,\dots,k_m}$ such that

$$\sum_{(k_1,\dots,k_m)\in\{1,\dots,n\}^m} c_{k_1,\dots,k_m}\beta_{k_1,\dots,k_m} = 0.$$

Fix $(p_1, \dots, p_m) \in \{1, \dots, n\}^m$. Then for each $(k_1, \dots, k_m) \in \{1, \dots, n\}^m$, we have

$$\beta_{k_1,\dots,k_m}(v_{p_1}, \dots, v_{p_m}) = \delta_{k_1 p_1} \cdots \delta_{k_m p_m}.$$

Evaluating the sum at $(v_{p_1}, \dots, v_{p_m})$ gives

$$\sum_{(k_1,\dots,k_m)\in\{1,\dots,n\}^m} c_{k_1,\dots,k_m}\beta_{k_1,\dots,k_m}(v_{p_1}, \dots, v_{p_m}) = 0,$$

which simplifies to $c_{p_1,\dots,p_m} = 0$. Since $(p_1, \dots, p_m)$ was arbitrary, it follows that all $c_{k_1,\dots,k_m} = 0$, proving that $\mathcal{B}$ is a linearly independent list of length n^m. Moreover, let $\beta \in V^{(m)}$ and $u_1, \dots, u_m \in V$ be arbitrary. By 3.114, we may write each u_ℓ as $u_\ell = \sum_{t=1}^{n} \phi_t(u_\ell) v_t$ for $\ell = 1, \dots, m$. Using the multilinearity of β and the fact that $\phi_{t_1}(u_1) \cdots \phi_{t_m}(u_m) = \beta_{t_1,\dots,t_m}(u_1, \dots, u_m)$, we have

$$\beta(u_1, \dots, u_m) = \sum_{t_1,\dots,t_m=1}^{n} \phi_{t_1}(u_1) \cdots \phi_{t_m}(u_m) \beta(v_{t_1}, \dots, v_{t_m}) = \sum_{t_1,\dots,t_m=1}^{n} \beta_{t_1,\dots,t_m}(u_1, \dots, u_m) \beta(v_{t_1}, \dots, v_{t_m}).$$

Since $u_1, \dots, u_m$ were arbitrary, this shows that

$$\beta = \sum_{t_1,\dots,t_m=1}^{n} \beta(v_{t_1}, \dots, v_{t_m}) \beta_{t_1,\dots,t_m}.$$

Hence, $\mathcal{B}$ spans $V^{(m)}$, and thus $\mathcal{B}$ is a basis of $V^{(m)}$ with $\dim V^{(m)} = n^m = (\dim V)^m$. □

Solution 9B.2 Observe that the map $(x_1, \dots, x_n) \mapsto x_i$ is linear for each $i = 1, \dots, n$, so each term in the definition of α is a product of three linear maps, one in each slot. Since each product is 3-linear and a linear combination of 3-linear maps is 3-linear, it follows that α is 3-linear. Let $x, y, z \in \mathbb{F}^n$ be arbitrary with $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, and $z = (z_1, \dots, z_n)$. Then

$$\begin{aligned}
\alpha(x, x, z) &= x_1x_2z_3 - x_2x_1z_3 - x_3x_2z_1 - x_1x_3z_2 + x_3x_1z_2 + x_2x_3z_1 = 0 \\
\alpha(x, y, x) &= x_1y_2x_3 - x_2y_1x_3 - x_3y_2x_1 - x_1y_3x_2 + x_3y_1x_2 + x_2y_3x_1 = 0 \\
\alpha(x, y, y) &= x_1y_2y_3 - x_2y_1y_3 - x_3y_2y_1 - x_1y_3y_2 + x_3y_1y_2 + x_2y_3y_1 = 0.
\end{aligned}$$

Hence, α is alternating. □

Solution 9B.3 We claim that an adjacent swap of two vectors in the arguments of α changes the sign of α. Fix some $j \in \{1, \dots, m-1\}$ and all slots except for the jth and $(j+1)$th slots. For arbitrary $u, w \in V$, consider $u + w$ in both the jth and $(j+1)$th slots. By hypothesis, $\alpha(\dots, u, u, \dots) = 0$ and $\alpha(\dots, w, w, \dots) = 0$, where the repeated entries are in the jth and $(j+1)$th slots. Expanding $\alpha(\dots, u + w, u + w, \dots)$ by linearity in the jth and $(j+1)$th slots, we get

$$0 = \alpha(\dots, u + w, u + w, \dots) = \alpha(\dots, u, w, \dots) + \alpha(\dots, w, u, \dots),$$

which shows that $\alpha(\dots, u, w, \dots) = -\alpha(\dots, w, u, \dots)$.
Now let $v_1, \dots, v_m \in V$ be arbitrary and suppose $v_\ell = v_k$ for some $\ell < k$. Apply the swap relation $k - \ell - 1$ times to move v_k to the $(\ell+1)$th slot, changing the sign each time. Then

$$\alpha(v_1, \dots, v_m) = (-1)^{k-\ell-1} \alpha(v_1, \dots, v_\ell, v_k, v_{\ell+1}, \dots, v_{k-1}, v_{k+1}, \dots, v_m).$$

The ℓth and $(\ell+1)$th slots of the right side now contain the same vector $v_\ell = v_k$, so by the hypothesis, the right side is 0. Hence, $\alpha(v_1, \dots, v_m) = 0$ whenever $v_\ell = v_k$ for some $\ell < k$, so α is alternating. □

Solution 9B.4 Take $V = \mathbb{F}^4$ and let U be the set in question. By 9.37, $\dim V^{(4)}_{\text{alt}} = 1$, so there exists a nonzero $\alpha \in V^{(4)}_{\text{alt}}$. Let $(x_1, x_2, x_3, x_4) \in V^4$ be such that $\alpha(x_1, x_2, x_3, x_4) \neq 0$. Applying 3.10 to the maps $v \mapsto \alpha(x_1, x_2, x_3, v)$ and $v \mapsto \alpha(v, 0, 0, x_4)$, we see that $\alpha(x_1, x_2, x_3, 0) = 0 = \alpha(0, 0, 0, x_4)$, so $(x_1, x_2, x_3, 0), (0, 0, 0, x_4) \in U$. However, $(x_1, x_2, x_3, 0) + (0, 0, 0, x_4) = (x_1, x_2, x_3, x_4) \notin U$. Hence, U is not closed under addition and therefore is not a subspace. □

Solution 9B.5 Suppose $v_\ell = v_k$ for some $\ell \neq k$. Let $\sigma = (j_1, \dots, j_m) \in \text{perm } m$, and let $\sigma' = (j'_1, \dots, j'_m)$ be σ with the entries equal to ℓ and k swapped. The map $\sigma \mapsto \sigma'$ is a bijection on perm m and $\text{sign}(\sigma') = -\text{sign}(\sigma)$ by 9.34. Because $v_\ell = v_k$, the lists $(v_{j_1}, \dots, v_{j_m})$ and $(v_{j'_1}, \dots, v_{j'_m})$ are equal. Then

$$\alpha(v_1, \dots, v_m) = \sum_{\sigma \in \text{perm } m} \text{sign}(\sigma')\beta(v_{j'_1}, \dots, v_{j'_m}) = \sum_{\sigma \in \text{perm } m} -\text{sign}(\sigma)\beta(v_{j_1}, \dots, v_{j_m}) = -\alpha(v_1, \dots, v_m).$$

Hence, $\alpha(v_1, \dots, v_m) = 0$ whenever $v_\ell = v_k$ for some $\ell \neq k$, so α is alternating. □

Solution 9B.6 Fix some $(k_1, \dots, k_m) \in \text{perm } m$ and $v_1, \dots, v_m \in V$. Let $u_p = v_{k_p}$ for $p = 1, \dots, m$. Distinctness of the entries in the lists $(j_1, \dots, j_m)$ and $(k_1, \dots, k_m)$ ensures that the map $(j_1, \dots, j_m) \mapsto (k_{j_1}, \dots, k_{j_m})$ lands in perm m. Since the k_p are distinct, this map is injective, and because perm m is finite, it is a bijection. By reindexing the sum using this bijection, we have

$$\alpha(u_1, \dots, u_m) = \sum_{(j_1, \dots, j_m) \in \text{perm } m} \beta(u_{j_1}, \dots, u_{j_m}) = \sum_{(j_1, \dots, j_m) \in \text{perm } m} \beta(v_{k_{j_1}}, \dots, v_{k_{j_m}}) = \alpha(v_1, \dots, v_m),$$

which gives the desired equality. □

Solution 9B.7

Axler Note. *This exercise shows that 9.39 can fail if the hypothesis that $n = \dim V$ is deleted.*

Let $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ be vectors in $\mathbb{R}^3$, let e_1, e_2, e_3 be the standard basis of $\mathbb{R}^3$, and define

$$\alpha(x, y) = x_1 y_2 - x_2 y_1.$$

With y fixed, $\alpha(x, y)$ is a linear function of x, and with x fixed, $\alpha(x, y)$ is a linear function of y, so α is bilinear. Moreover, $\alpha(x, x) = x_1 x_2 - x_2 x_1 = 0$ for all $x \in \mathbb{R}^3$, so α is alternating. Because $\alpha(e_1, e_2) = 1 \cdot 1 - 0 \cdot 0 = 1 \neq 0$, we see that α is nonzero. Finally, $v_1 = e_1$ and $v_2 = e_3$ form a linearly independent list in $\mathbb{R}^3$, and $\alpha(v_1, v_2) = \alpha(e_1, e_3) = 0$. □

§9C

Solution 9C.1 This statement is false in general. Consider $S = T = I$ with $\dim V = 2$. For any $\alpha \in V_{\text{alt}}^{(2)}$, we have $\alpha_I = \alpha$, and

$$\alpha_{2I}(u, v) = \alpha(2u, 2v) = 4\alpha(u, v).$$

Hence, $\det(S + T) = \det(2I) = 4 \neq 1 + 1 = \det S + \det T$. □

Solution 9C.2 Let A be n-by-n with $n > 1$ so that B is well-defined. Let $\Lambda \subseteq \text{perm } n$ denote the set of permutations with 1 in the first position. In 9.46, the factor $A_{j_1,1} = 0$ unless $j_1 = 1$. Therefore, the only permutations that contribute to the sum are of the form $\sigma = (1, j_2, \ldots, j_n)$, so

$$\det A = A_{1,1} \sum_{\sigma \in \Lambda} (\operatorname{sign} \sigma) \prod_{k=2}^{n} A_{j_k,k}.$$

Let $\ell_p = j_{p+1} - 1$ for $1 \leq p \leq n-1$, and let $\tau = (\ell_1, \ldots, \ell_{n-1})$. Then the map $\sigma \mapsto \tau$ is a bijection from Λ to perm $(n-1)$. By 9.32, no pair $(1, j_p)$ with $p \geq 2$ contributes to the sign. Moreover, for $2 \leq q < r$, the value q appears after r in σ if and only if $q - 1$ appears after $r - 1$ in τ, which holds because τ is σ with the first element removed and each remaining element decreased by 1. Therefore, $\operatorname{sign} \sigma = \operatorname{sign} \tau$. Because $B_{p,q} = A_{p+1,q+1}$ for $1 \leq p, q \leq n-1$, we have

$$\prod_{q=1}^{n-1} B_{\ell_q,q} = \prod_{k=2}^{n} A_{j_k,k}.$$

Hence,

$$\det A = A_{1,1} \sum_{\tau \in \text{perm}\,(n-1)} (\operatorname{sign} \tau) \prod_{q=1}^{n-1} B_{\ell_q,q} = A_{1,1} \det B.$$

□

Solution 9C.3 By 8.18, V has a basis with respect to which $\text{Mat}(T)$ is strictly upper triangular. By 3.35, $\text{Mat}(I + T) = I + \text{Mat}(T)$ is upper triangular with all diagonal entries equal to 1. By 9.53 and 9.48, $\det(I + T) = 1$. □

Solution 9C.4 Let $s_1 \geq \cdots \geq s_n$ denote the singular values of S. By 7.88(a) and 9.60, $\|S\| = s_1$ and $|\det S| = s_1 \cdots s_n$.

$\Rightarrow$ Suppose S is unitary. Then S is an isometry, so 7.69 gives $s_1 = \cdots = s_n = 1$, and hence $|\det S| = s_1 \cdots s_n = 1$ and $\|S\| = s_1 = 1$.

$\Leftarrow$ Suppose $|\det S| = \|S\| = 1$. Then $s_1 = \|S\| = 1$ and $s_1 \cdots s_n = |\det S| = 1$. Because $0 \leq s_k \leq s_1 = 1$ for all k, we must have $s_k = 1$ for all k. Since $\det S \neq 0$, 9.50 yields that S is invertible. Therefore, S is an invertible isometry by 7.69, and hence unitary. □

Solution 9C.5 Proceed by induction on m. The base case $m = 1$ is trivial, since $\det A = \det A_1$. Suppose $m > 1$ and the result holds for block upper-triangular matrices with $m - 1$ diagonal blocks. Write A as

$$A = \begin{pmatrix} A_1 & * \\ 0 & B \end{pmatrix},$$

where B is block upper triangular with diagonal blocks $A_2, \ldots, A_m$. By the induction hypothesis, it suffices to show that $\det A = (\det A_1)(\det B)$. Let A_1 be n-by-n, B be p-by-p, and let $e_1, \ldots, e_n$ denote the standard basis of $\mathbb{F}^n$. For $v_1, \ldots, v_n \in \mathbb{F}^n$, let $\alpha(v_1, \ldots, v_n)$ be the determinant of A with the first n columns replaced by $(v_1, 0), \ldots, (v_n, 0)$, where 0 denotes the zero vector in $\mathbb{F}^p$. Since the map $v \mapsto (v, 0)$ is linear from $\mathbb{F}^n$ to $\mathbb{F}^{n+p}$, 9.45 implies that α is an alternating n-linear form on $\mathbb{F}^n$. By 9.36 applied to the standard basis and 9.46, we have

$$\alpha(v_1, \ldots, v_n) = \alpha(e_1, \ldots, e_n) \det(v_1, \ldots, v_n).$$

Observe that

$$\alpha(e_1, \ldots, e_n) = \det \begin{pmatrix} I & * \\ 0 & B \end{pmatrix} = \det B,$$

where Exercise 9C2 is applied n times to expand the determinant along the first n columns. By taking $v_1, \ldots, v_n$ to be the columns of A_1, we get $\det A = (\det A_1)(\det B)$, as desired. □

Solution 9C.6 By 9.45, $\alpha(u_1, \ldots, u_n) = \det \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix}$ is an alternating n-linear form on $\mathbb{F}^n$. Applying 9.35 to α and the list $v_1, \ldots, v_n$, we get

$$\alpha(v_{m_1}, \ldots, v_{m_n}) = \operatorname{sign}(m_1, \ldots, m_n)\alpha(v_1, \ldots, v_n),$$

where $\alpha(v_1, \ldots, v_n) = \det A$, which completes the proof. □

Solution 9C.7 Let $n = \dim V$. By 9.65, $p(0) = (-1)^n \det T$, which is nonzero by 9.50. Note that for $z \neq 0$,

$$zI - T^{-1} = (-zI)T^{-1}\left(\frac{1}{z}I - T\right).$$

By 9.49(a), 9.53, 9.48, and 9.50, we get

$$\begin{aligned} q(z) &= \det\left((-zI)T^{-1}\left(\frac{1}{z}I - T\right)\right) \\ &= \det(-zI)\det(T^{-1})\det\left(\frac{1}{z}I - T\right) \\ &= (-z)^n \det(T^{-1})\, p\left(\frac{1}{z}\right) \\ &= \frac{z^n}{(-1)^n \det T}\, p\left(\frac{1}{z}\right) \\ &= \frac{z^n}{p(0)}\, p\left(\frac{1}{z}\right). \end{aligned}$$

□

Solution 9C.8 Let $n = \dim V$ and let p be the characteristic polynomial of T. By 9.65, p is monic with $\deg p = n$ and $p(0) = (-1)^n \det T$. By 9.51, p has no real zeros, so 4.16 yields that p has no real linear factors, and

$$p(z) = \prod_{j=1}^{M} (z^2 + b_j z + c_j),$$

where $b_j, c_j \in \mathbb{R}$ and $b_j^2 < 4c_j$ for each j. Comparing degrees gives $n = 2M$, and since $b_j^2 < 4c_j$, we have $c_j > 0$ for each j, so

$$\det T = p(0) = \prod_{j=1}^{M} c_j > 0. \qquad \square$$

Solution 9C.9 Let $n = \dim V$ and let p be the characteristic polynomial of T. By 9.65, p is monic with $\deg p = n$ and $p(0) = (-1)^n \det T$. By 4.16, we may write

$$p(z) = \prod_{j=1}^{m} (z - \lambda_j) \prod_{k=1}^{M} (z^2 + b_k z + c_k),$$

where $\lambda_j \in \mathbb{R}$, $b_k, c_k \in \mathbb{R}$, and $c_k > b_k^2/4 \geq 0$ for each k. Since $\deg p$ is even and the second product has degree $2M$, it follows that m must also be even. Evaluating p at 0 gives

$$p(0) = \prod_{j=1}^{m} (-\lambda_j) \prod_{k=1}^{M} c_k < 0.$$

By 9.51, each λ_j is an eigenvalue of T. Suppose T has at most one eigenvalue. Then either $m = 0$, in which case $p(0) = \prod_{k=1}^{M} c_k \geq 0$, or $m \geq 2$, in which case all λ_j are equal to some λ, and $p(0) = \prod_{j=1}^{m} (-\lambda) \prod_{k=1}^{M} c_k \geq 0$. In either case, we get a contradiction with $p(0) < 0$, so T must have at least two distinct eigenvalues. $\square$

Solution 9C.10 Let $n = \dim V$ and let p be the characteristic polynomial of T. By 9.65, p is monic with $\deg p = n$. By 4.16, we may write

$$p(z) = \prod_{j=1}^{m} (z - \lambda_j) \prod_{k=1}^{M} (z^2 + b_k z + c_k),$$

where $\lambda_j \in \mathbb{R}$, $b_k, c_k \in \mathbb{R}$, and $b_k^2 < 4c_k$ for each k. Comparing degrees, we have $n = m + 2M$, which forces $m \geq 1$ since n is odd. Then $p(\lambda_1) = 0$, so λ_1 is an eigenvalue of T by 9.51. $\square$

Solution 9C.11

Axler Note. *If $\mathbb{F} = \mathbb{C}$, $T \in \text{End}(V)$, and $\det T \neq 0$, then T has a square root (see 8.41).*

This is false. Suppose V is a real vector space with $\dim V = 2$ and basis v_1, v_2. By the linear map lemma, define $T \in \text{End}(V)$ by $Tv_1 = -v_1$ and $Tv_2 = -2v_2$. Then

Mat(T) with respect to the basis v_1, v_2 is diagonal with diagonal entries -1 and -2, so $\det T = 2 > 0$ by 9.53 and 9.48. Moreover, for any $\lambda, \mu \in \mathbb{R}$, we have

$$(T + I)(\lambda v_1 + \mu v_2) = \lambda(-v_1 + v_1) + \mu(-2v_2 + v_2) = -\mu v_2,$$

so $E(-1, T) = \ker(T + I) = \operatorname{span} v_1$.

Suppose T has a square root S, so $S^2 = T$. Then $ST = S^3 = TS$, so $E(-1, T)$ is invariant under S by 5.75, and $Sv_1 = \nu v_1$ for some $\nu \in \mathbb{R}$. However, $-v_1 = Tv_1 = S^2 v_1 = S(\nu v_1) = \nu S v_1 = \nu^2 v_1$ gives $v_1 = 0$ or $\nu^2 = -1$. Since v_1 is a basis vector and $\nu \in \mathbb{R}$, this is impossible. Therefore, T cannot have a square root. □

Solution 9C.12 Note that $zS - T = S(zI - S^{-1}T)$, so by 9.49(a), $p(z) = (\det S)q(z)$, where $q(z)$ is the characteristic polynomial of $S^{-1}T$. By 9.65, q is monic with $\deg q = \dim V$. By 9.50, $\det S \neq 0$, so p is a polynomial of degree $\dim V$ and the coefficient of $z^{\dim V}$ in p is $\det S$. □

Solution 9C.13 By 9.62, the characteristic polynomial of T is

$$\det(zI - T) = \prod_{k=1}^{n} (z - \lambda_k).$$

In general, the coefficient of z^{n-k} in the characteristic polynomial of T is $(-1)^k$ times the sum of all products of k distinct eigenvalues of T.

(a) With $r = 2$, the coefficient of z^{n-2} is

$$(-1)^2 \sum_{1 \le j < k \le n} \lambda_j \lambda_k = \sum_{1 \le i < j \le n} \lambda_i \lambda_j.$$

(b) With $r = n - 1$, the coefficient of z is

$$(-1)^{n-1} \sum_{1 \le j_1 < \cdots < j_{n-1} \le n} \lambda_{j_1} \cdots \lambda_{j_{n-1}}.$$ □

Solution 9C.14 By 7.39, $(\sqrt{T})^2 = T$, so $\det T = (\det \sqrt{T})^2$ by 9.49(a). Since $\sqrt{T}$ is positive, $\det \sqrt{T} \ge 0$ by 9.59, so taking square roots gives $\det \sqrt{T} = \sqrt{\det T}$. □

Solution 9C.15 By the polar decomposition, $T = S\sqrt{T^*T}$ for some unitary operator S, where T^*T is positive by 7.64(a). By 9.58 and 9.59, $|\det S| = 1$ and $\det \sqrt{T^*T} \ge 0$. Therefore, 9.49(a) gives

$$|\det T| = |\det S|\,|\det \sqrt{T^*T}| = \det \sqrt{T^*T},$$

and Exercise 9C14 gives $\det \sqrt{T^*T} = \sqrt{\det(T^*T)}$. □

Solution 9C.16 Let $n = \dim V$. By 9.65 and 8.56 applied to $-T$,

$$\det(zI + T) = z^n + (\operatorname{tr} T)z^{n-1} + \sum_{k=0}^{n-2} \lambda_k z^k$$

for $\lambda_0, \ldots, \lambda_{n-2} \in \mathbb{F}$. For nonzero x, we have $I + xT = x(\frac{1}{x}I + T)$. By 9.53, 3.38, and 9.57(c) applied to each column of $x \operatorname{Mat}(\frac{1}{x}I + T)$, we have $\det(x(\frac{1}{x}I + T)) = x^n \det(\frac{1}{x}I + T)$. Using the substitution $z = \frac{1}{x}$ and the fact that $g(0) = \det I = 1$ by 9.53 and 9.48, we have for all $x \in \mathbb{F}$ that

$$g(x) = 1 + (\operatorname{tr} T)x + \sum_{k=2}^{n} \lambda_{n-k} x^k.$$

Since each exponent in the sum is at least 2, evaluating $(x^k)'$ at $x = 0$ gives 0 for each $k \geq 2$, so $g'(0) = \operatorname{tr} T$. □

Solution 9C.17 Let E be the ellipsoid in question. Let Ω be the unit ball in $\mathbb{R}^3$ with $\text{volume}(\Omega) = \frac{4}{3}\pi$. Define $T \in \text{End}(\mathbb{R}^3)$ by $T(x, y, z) = (ax, by, cz)$. We claim that $T(\Omega) = E$.

$\subseteq$ Suppose $(x, y, z) \in \Omega$. Then $x^2 + y^2 + z^2 < 1$, so

$$\frac{(ax)^2}{a^2} + \frac{(by)^2}{b^2} + \frac{(cz)^2}{c^2} = x^2 + y^2 + z^2 < 1.$$

Hence, $(ax, by, cz) \in E$, so $T(\Omega) \subseteq E$.

$\supseteq$ Suppose $(p, q, r) \in E$. Then

$$\frac{p^2}{a^2} + \frac{q^2}{b^2} + \frac{r^2}{c^2} = \left(\frac{p}{a}\right)^2 + \left(\frac{q}{b}\right)^2 + \left(\frac{r}{c}\right)^2 < 1,$$

so $(\frac{p}{a}, \frac{q}{b}, \frac{r}{c}) \in \Omega$, and thus $T(\frac{p}{a}, \frac{q}{b}, \frac{r}{c}) = (p, q, r) \in T(\Omega)$, whence $E \subseteq T(\Omega)$.

Therefore, $T(\Omega) = E$. With respect to the standard basis of $\mathbb{R}^3$, the matrix of T is diagonal with diagonal entries a, b, and c, so $\det T = abc$ by 9.53 and 9.48. By 9.61, $\text{volume } E = |\det T|(\text{volume } \Omega) = \frac{4}{3}abc\pi$. □

Solution 9C.18 Since A is invertible, $Ax = 0$ only when $x = 0$, so the columns $v_1, \ldots, v_n$ of A are linearly independent by 3.50. By QR factorization, we may write $A = QR$, where Q is unitary and R is upper triangular with positive diagonal entries. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$ and let $T \in \text{End}(\mathbb{F}^n)$ be such that $\operatorname{Mat}(T, (e_1, \ldots, e_n)) = Q$. By 7.57(d), $Q^*Q = I$, where $(Q^*Q)_{j,k} = \langle Q_{\cdot,k}, Q_{\cdot,j} \rangle$, so the columns of Q are orthonormal, and hence form an orthonormal basis of $\mathbb{F}^n$ by 6.28. By 3.75, $Te_k = Q_{\cdot,k}$ for each $k = 1, \ldots, n$, so T is unitary by 7.53(d). By 9.43 and 9.58, $|\det Q| = |\det T| = 1$. By 9.49(b), 9.48, and $R_{k,k} > 0$ for each k, we have

$$|\det A| = |\det Q|\,|\det R| = \prod_{k=1}^{n} R_{k,k}.$$

Since $v_k = QR_{\cdot,k}$ by 3.48, we have $\|v_k\| = \|R_{\cdot,k}\|$ by 7.57(c). Because R is upper triangular,

$$\|R_{\cdot,k}\|^2 = R_{k,k}^2 + \sum_{j<k} |R_{j,k}|^2.$$

Since $0 < R_{k,k} \leq \|R_{\cdot,k}\|$ for each k, Hadamard's inequality for A can be read as

$$|\det A| = \prod_{k=1}^{n} R_{k,k} \leq \prod_{k=1}^{n} \|R_{\cdot,k}\| = \prod_{k=1}^{n} \|v_k\|.$$

The products are equal if and only if $\|R_{\cdot,k}\| = R_{k,k}$ for each $k = 1, \ldots, n$, i.e., $R_{j,k} = 0$ for all k and $j < k$ if and only if R is diagonal.
$\Rightarrow$ Suppose equality holds in Hadamard's inequality for A. Then R is diagonal, so $v_k = R_{k,k} Q_{\cdot,k}$ for each $k = 1, \ldots, n$. Since the columns of Q are orthonormal, $\langle v_j, v_k \rangle = 0$ for all $j \neq k$, i.e., each column of A is orthogonal to the other columns.

$\Leftarrow$ Suppose the columns $v_1, \ldots, v_n$ of A are mutually orthogonal, and let $e_1, \ldots, e_n$ be the standard orthonormal basis of $\mathbb{F}^n$. Let $T \in \operatorname{End}(\mathbb{F}^n)$ be such that $\operatorname{Mat}(T, (e_1, \ldots, e_n)) = A$. By 7.9 and 3.43, $\operatorname{Mat}(T^*T) = A^*A$. By 9.43, 9.53, and 9.60, $|\det A| = |\det T| = \sqrt{\det(T^*T)} = \sqrt{\det(A^*A)}$. By 3.46, the (j,k)th entry of A^*A is $\langle v_k, v_j \rangle$. By orthogonality of the columns, the off-diagonal entries of A^*A are zero, and the diagonal entries are $\langle v_k, v_k \rangle = \|v_k\|^2$. By 9.48,

$$\det(A^*A) = \prod_{k=1}^{n} \|v_k\|^2.$$

Taking square roots yields the desired result. □

Solution 9C.19 Let $A = \operatorname{Mat}(\sqrt{T}, (e_1, \ldots, e_n))$. Since $\sqrt{T}$ is positive, it is self-adjoint. By 3.75, $A_{\cdot,k} = \operatorname{Mat}(\sqrt{T} e_k)$ for each $k = 1, \ldots, n$. By 6.30(c) and self-adjointness,

$$\langle A_{\cdot,k}, A_{\cdot,j} \rangle = \langle \sqrt{T} e_k, \sqrt{T} e_j \rangle = \langle T e_k, e_j \rangle,$$

and therefore, $\|A_{\cdot,k}\|^2 = \langle T e_k, e_k \rangle$ for each $k = 1, \ldots, n$. By 9.53 and Exercise 9C14, $\det A = \det \sqrt{T} = \sqrt{\det T} \geq 0$, so $\det T = |\det A|^2$.

(a) By Hadamard's inequality,

$$\det T = |\det A|^2 \leq \prod_{k=1}^{n} \|A_{\cdot,k}\|^2 = \prod_{k=1}^{n} \langle T e_k, e_k \rangle.$$

(b) $\Rightarrow$ Suppose the inequality in (a) is an equality. Since T is invertible, $\det T \neq 0$, so $\det A = \sqrt{\det T} \neq 0$, and thus $\sqrt{T}$ is invertible by 9.50, whence A is also invertible by 3.86. Taking nonnegative square roots, equality holds in Hadamard's inequality for A. Thus, $\langle T e_k, e_j \rangle = 0$ for all $j \neq k$ by Exercise 9C18. By 6.30(a), $T e_k = \langle T e_k, e_k \rangle e_k$ for each $k = 1, \ldots, n$, so e_k is an eigenvector of T.

$\Leftarrow$ Suppose e_k is an eigenvector of T for each $k = 1, \ldots, n$. Then $T e_k = \lambda_k e_k$ for some $\lambda_k \in \mathbb{F}$ and each $k = 1, \ldots, n$. It follows that $\operatorname{Mat}(T, (e_1, \ldots, e_n))$ is diagonal with diagonal entries $\lambda_k = \langle T e_k, e_k \rangle$ for each $k = 1, \ldots, n$. By 9.53 and 9.48,

$$\det T = \prod_{k=1}^{n} \lambda_k = \prod_{k=1}^{n} \langle T e_k, e_k \rangle.$$ □

Solution 9C.20 Note that $c \geq |A_{1,1}| \geq 0$. For each $k = 1, \ldots, n$, the column v_k of A satisfies

$$\|v_k\|^2 = \sum_{j=1}^{n} |A_{j,k}|^2 \leq c^2 n.$$

By Hadamard's inequality,

$$|\det A| \leq \prod_{k=1}^{n} \|v_k\| \leq \prod_{k=1}^{n} \sqrt{c^2 n} = c^n n^{n/2}. \qquad \square$$

Solution 9C.21

Axler Note. *Recall that $\mathbb{C}^{n,n}$ denotes the set of n-by-n matrices with entries in $\mathbb{C}$. This exercise shows that the determinant is the unique function defined on square matrices that is multiplicative and has the desired behavior on diagonal matrices. This result is analogous to Exercise 8D10, which shows that the trace is uniquely determined by its algebraic properties.*

Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{C}^n$. Since I is diagonal, $\delta(I) = 1$. For every invertible matrix $P \in \mathbb{C}^{n,n}$, we have $\delta(P)\delta(P^{-1}) = \delta(PP^{-1}) = \delta(I) = 1$, so $\delta(P^{-1}XP) = \delta(X)$ for all $X \in \mathbb{C}^{n,n}$. Note that $\det I = 1$ by 9.48, so by 9.49(b), $\det(P^{-1}XP) = \det(P^{-1})(\det X)(\det P) = \det X$ for all $X \in \mathbb{C}^{n,n}$ and invertible P. Therefore, $\delta(X) = \det X$ whenever $X = P^{-1}DP$ for some invertible P and diagonal matrix D. We split by whether X is invertible or not.

- Suppose X is invertible. By 5.47 and 3.84, $X = P^{-1}YP$ for some invertible P and upper-triangular matrix Y, where Y has diagonal entries $\lambda_1, \ldots, \lambda_n$. Since $\det Y = \det X \neq 0$ by 9.48 and 9.50, $\lambda_k \neq 0$ for all k. Let D be diagonal with diagonal entries λ_k / k for each $k = 1, \ldots, n$. Then $Y' = D^{-1}Y$ is upper triangular with entries $1, \ldots, n$ on the diagonal. By 5.41, the operator $T \in \mathrm{End}(\mathbb{C}^n)$ such that $\mathrm{Mat}(T, (e_1, \ldots, e_n)) = Y'$ has eigenvalues $1, \ldots, n$. By 5.58, T is diagonalizable, so 3.84 gives $Y' = Q^{-1}D'Q$ for invertible Q and diagonal D'. By the beginning calculation, $\det Y' = \det D'$, so $\det Y = (\det D)(\det Y') = (\det D)(\det D')$ by 9.49(b). Therefore,

$$\delta(X) = \delta(P^{-1}YP) = \delta(Y) = \delta(D)\delta(Y') = (\det D)(\det D') = \det Y = \det X.$$

- Suppose X is not invertible. Let $S \in \mathrm{End}(\mathbb{C}^n)$ be such that $\mathrm{Mat}(S, (e_1, \ldots, e_n)) = X$. By 3.86, S is not invertible, so it is not injective by 3.65. Then there exists nonzero $v \in \mathbb{C}^n$ such that $Sv = 0$, whence $Xv = \mathrm{Mat}(Sv) = 0$ by 3.76. By 2.32, extend v to a basis $v, v_2, \ldots, v_n$ of $\mathbb{C}^n$, and let Q be the matrix whose columns are $v, v_2, \ldots, v_n$. Since $Q = \mathrm{Mat}(I, (v, v_2, \ldots, v_n), (e_1, e_2, \ldots, e_n))$, 3.82 gives that Q is invertible. Let $M = Q^{-1}XQ$. Then M has the first column equal to 0, and thus $M = ME$, where E is diagonal with diagonal entries $0, 1, \ldots, 1$. By 9.53 and 9.50, we have $\det X = \det S = 0$. Therefore,

$$\delta(X) = \delta(M) = \delta(M)\delta(E) = \delta(M) \cdot 0 = 0 = \det X.$$

$\square$

§9D

Solution 9D.1 $\Rightarrow$ We proceed by contrapositive. Suppose $v \neq 0$ and $w \neq 0$ and let $n = \dim V$. Then the list v is linearly independent, so it extends to a basis $v, v_2, \ldots, v_n$ of V by 2.32. Let $\phi \in V'$ be the first element of the dual basis of $v, v_2, \ldots, v_n$, so $\phi(v) = 1$. A similar argument applies to w to obtain $\psi \in W'$ such that $\psi(w) = 1$. Then

$$(v \otimes w)(\phi, \psi) = \phi(v)\psi(w) = 1 \neq 0,$$

so $v \otimes w \neq 0$.

$\Leftarrow$ Suppose $v = 0$ or $w = 0$. Then for all $\phi \in V'$ and $\psi \in W'$, we have

$$(v \otimes w)(\phi, \psi) = \phi(v)\psi(w) = 0,$$

and so $v \otimes w = 0$. □

Solution 9D.2 Let e_1, e_2, e_3 be the standard basis of $\mathbb{R}^3$. Define the vectors

$$\begin{aligned} v_1 &= e_1, & v_2 &= e_2, & v_3 &= e_1 + 2e_2, \\ w_1 &= e_3, & w_2 &= 2e_3, & w_3 &= -e_3. \end{aligned}$$

Let $x = e_1 \otimes e_3$ and $y = e_2 \otimes e_3$. By 9.73, $v_1 \otimes w_1 = x$, $v_2 \otimes w_2 = 2y$, and $v_3 \otimes w_3 = -x - 2y$, which sum to 0. By 9.74(a), the list x, y is linearly independent. Note that $\lambda x + \mu(-x - 2y) = (\lambda - \mu)x - 2\mu y$, and $\lambda(2y) + \mu(-x - 2y) = -\mu x + (2\lambda - 2\mu)y$ for $\lambda, \mu \in \mathbb{R}$. Since x and y are linearly independent, these expressions can only be zero if $\lambda = \mu = 0$. Therefore, no two of $x, 2y, -x - 2y$ are scalar multiples of each other. □

Solution 9D.3

Appendix. *An alternate proof is given at the end of this solution.*

Let $n = \dim W$ and let $u_1, \ldots, u_n$ be a basis of W. Write $w_k = \sum_{j=1}^{n} \lambda_{j,k} u_j$ for $\lambda_{j,k} \in \mathbb{F}$. By 9.73, we have

$$0 = \sum_{k=1}^{m} v_k \otimes w_k = \sum_{k=1}^{m} v_k \otimes \sum_{j=1}^{n} \lambda_{j,k} u_j = \sum_{k=1}^{m} \sum_{j=1}^{n} \lambda_{j,k} (v_k \otimes u_j).$$

By 9.74(a), the list $\{v_k \otimes u_j\}_{k=1,\ldots,m;\, j=1,\ldots,n}$ is linearly independent. Therefore, $\lambda_{j,k} = 0$ for all j and k, and hence $w_k = 0$ for all k.

Alternate proof. The proof below uses at dual basis functionals rather than relying on expanding in a basis.

Alternate proof. Civil Service Pigeon. Since $v_1, \ldots, v_m$ is linearly independent, there exist linear functionals $\varphi_1, \ldots, \varphi_m \in V'$ such that $\varphi_j(v_k) = \delta_{j,k}$. For any $j \in \{1, \ldots, m\}$ and any arbitrary $\tau \in W'$,

$$0 = \left(\sum_{k=1}^{m} v_k \otimes w_k \right)(\varphi_j, \tau) = \sum_{k=1}^{m} \varphi_j(v_k)\tau(w_k) = \tau(w_j).$$

Since $\tau(w_j) = 0$ for all $\tau \in W'$, $w_j = 0$ for each j. □

Solution 9D.4

Appendix. *An alternate proof is given at the end of this solution.*

Let $n = \dim V$ and $m = \dim W$. Fix bases $v_1, \dots, v_n$ of V and $w_1, \dots, w_m$ of W. Let $u = v_1 \otimes w_1 + v_2 \otimes w_2$. We claim that $u \neq v \otimes w$ for all $(v, w) \in V \times W$. To see this, write $v = \sum_{k=1}^{n} \lambda_k v_k$. By 9.73,

$$v \otimes w - u = \sum_{k=1}^{n} v_k \otimes x_k, \quad x_k = \begin{cases} \lambda_k w - w_k, & k = 1, 2 \\ \lambda_k w, & k > 2. \end{cases}$$

By 2.22, at least one of w_1 or w_2 is not in $\operatorname{span}(w)$. If $w_k \notin \operatorname{span}(w)$, then $x_k \neq 0$, so $v \otimes w - u \neq 0$ by Exercise 9D3. Therefore, the set contains $v_1 \otimes w_1$ and $v_2 \otimes w_2$, but not their sum u, so it is not closed under addition and hence not a subspace of $V \otimes W$.

Alternate proof. Similar to the alternate proof of Exercise 9D3.

Alternate proof. Civil Service Pigeon. Since $\dim V > 1$ and $\dim W > 1$, choose linearly independent $v_1, v_2 \in V$ and linearly independent $w_1, w_2 \in W$. For the sake of contradiction, suppose that $v_1 \otimes w_1 + v_2 \otimes w_2 = v \otimes w$ for some $v \in V$ and $w \in W$. Choose linear functionals $\varphi, \phi \in V'$ such that $\varphi(v_1) = 1, \varphi(v_2) = 0$ and $\phi(v_1) = 0, \phi(v_2) = 1$.

Applying the map $V \otimes W \to W$ given by $x \otimes y \mapsto f(x)y$ for each functional yields

$$w_1 = \varphi(v)w, \qquad w_2 = \phi(v)w.$$

Hence, w_1 and w_2 are both scalar multiples of w, contradicting linear independence. Thus, the set is not a subspace of $V \otimes W$. □

Solution 9D.5 $\subseteq$ Note that the kth column of $v \otimes w$ is $w_k v$. By 2.6, the span of the columns is a subspace of $\operatorname{span}(v)$, so the rank of $v \otimes w$ is at most $\dim \operatorname{span}(v)$ by 2.37. Since $\dim \operatorname{span}(v) \leq 1$ by 2.30, the rank of $v \otimes w$ is at most one.

$\supseteq$ Suppose A is an m-by-n matrix of rank at most one. If $A = 0$, then $A = 0 \otimes 0$. Otherwise, $\operatorname{rank} A = 1$, so by 3.56, $A = BC$ for $B \in \mathbb{F}^{m,1}$ and $C \in \mathbb{F}^{1,n}$. Let $x = (B_{1,1}, \dots, B_{m,1})$ and $y = (C_{1,1}, \dots, C_{1,n})$. By 3.46, $A_{j,k} = B_{j,1} C_{1,k}$ for all j and k, so $A = x \otimes y$ by 9.76. □

Solution 9D.6 Identify each element of $\mathbb{F}^m \otimes \mathbb{F}^n$ with its coefficient matrix with respect to the basis of 9.74(b). Note that this identification is linear and extends 9.76. We claim that the set

$$\{v_1 \otimes w_1 + v_2 \otimes w_2 : v_1, v_2 \in \mathbb{F}^m, w_1, w_2 \in \mathbb{F}^n\}$$

is the set of m-by-n matrices that have rank at most two.

$\subseteq$ Let $w_{p,q}$ denote the qth entry of w_p for $p = 1, 2$ and $q = 1, \dots, n$, and let $u = v_1 \otimes w_1 + v_2 \otimes w_2$. Note that the kth column of u is $w_{1,k} v_1 + w_{2,k} v_2$. By 2.6, the span of the columns is a subspace of $\operatorname{span}(v_1, v_2)$, so the rank of u is at most $\dim \operatorname{span}(v_1, v_2)$ by 2.37. Since $\dim \operatorname{span}(v_1, v_2) \leq 2$ by 2.30, the rank of u is at most two.

$\supseteq$ Suppose A is an m-by-n matrix of rank at most two. If A has rank at most one, then Exercise 9D5 and Exercise 9D1 imply that $A = v \otimes w + 0 \otimes 0$ for some $v \in \mathbb{F}^m$ and $w \in \mathbb{F}^n$. If A has rank two, then by 3.56, $A = BC$ for some $B \in \mathbb{F}^{m,2}$

and $C \in \mathbb{F}^{2,n}$. Let x_ℓ be the ℓth column of B and y_ℓ be the ℓth row of C for $\ell = 1, 2$. By 3.46, $A_{j,k} = B_{j,1}C_{1,k} + B_{j,2}C_{2,k}$ for all j and k, so $A = x_1 \otimes y_1 + x_2 \otimes y_2$. □

Solution 9D.7

Appendix. *An alternate proof is given at the end of this solution.*

Let $n = \dim V$ and $m = \dim W$. Fix bases $e_1, \dots, e_n$ of V and $f_1, \dots, f_m$ of W. Let $u = e_1 \otimes f_1 + e_2 \otimes f_2 + e_3 \otimes f_3$. As in Exercise 9D4, we claim that $u \neq v_1 \otimes w_1 + v_2 \otimes w_2$ for any $v_1, v_2 \in V$ and $w_1, w_2 \in W$. Indeed, write $v_j = \sum_{k=1}^{n} \lambda_{j,k} e_k$. By 9.73,

$$v_1 \otimes w_1 + v_2 \otimes w_2 - u = \sum_{k=1}^{n} e_k \otimes x_k, \quad x_k = \begin{cases} \lambda_{1,k} w_1 + \lambda_{2,k} w_2 - f_k & \text{for } k \leq 3, \\ \lambda_{1,k} w_1 + \lambda_{2,k} w_2 & \text{for } k > 3. \end{cases}$$

By 2.22, not all of f_1, f_2, f_3 lie in $\operatorname{span}(w_1, w_2)$. Choose $k \leq 3$ such that $f_k \notin \operatorname{span}(w_1, w_2)$. Since $x_k + f_k \in \operatorname{span}(w_1, w_2)$, if $x_k = 0$, then $f_k \in \operatorname{span}(w_1, w_2)$, which is a contradiction, so $x_k \neq 0$. Hence, $v_1 \otimes w_1 + v_2 \otimes w_2 - u \neq 0$ by Exercise 9D3, and $u \notin \{v_1 \otimes w_1 + v_2 \otimes w_2 : v_1, v_2 \in V \text{ and } w_1, w_2 \in W\}$.

Alternate proof. This proof gives tensors as linear maps from V' to W, then compares ranks.

Alternate proof. Civil Service Pigeon. Since $\dim V > 2$ and $\dim W > 2$, choose linearly independent vectors $e_1, e_2, e_3 \in V$ and $f_1, f_2, f_3 \in W$. Under the natural isomorphism between $V \otimes W$ and $\mathcal{L}(V', W)$, $v \otimes w$ corresponds to a linear map of rank at most 1, implying any element of the form $v_1 \otimes w_1 + v_2 \otimes w_2$ corresponds to a map of rank at most 2. However, $u = e_1 \otimes f_1 + e_2 \otimes f_2 + e_3 \otimes f_3 \in V \otimes W$ corresponds to a linear map of rank 3. Therefore, u cannot be written as the sum of two pure tensors, proving the set is not all of $V \otimes W$. □

Solution 9D.8 By 9.79(a), there exists a linear map $\hat{\Gamma}: V \otimes W \to U$ such that $\hat{\Gamma}(v \otimes w) = \Gamma(v, w)$ for all $v \in V$ and $w \in W$. By linearity and 3.10,

$$\sum_{k=1}^{m} \Gamma(v_k, w_k) = \sum_{k=1}^{m} \hat{\Gamma}(v_k \otimes w_k) = \hat{\Gamma}\left(\sum_{k=1}^{m} v_k \otimes w_k\right) = \hat{\Gamma}(0) = 0. \qquad \square$$

Solution 9D.9 Note that the map $\Gamma: V \times W \to V \otimes W$ given by $\Gamma(v, w) = Sv \otimes Tw$ is bilinear, since each slot is linear by linearity of S and T and 9.73. By 9.79(a) with $U = V \otimes W$, there exists a unique linear map $\hat{\Gamma}: V \otimes W \to V \otimes W$ such that $\hat{\Gamma}(v \otimes w) = \Gamma(v, w) = Sv \otimes Tw$ for all $v \in V$ and $w \in W$. □

Solution 9D.10 $\Rightarrow$ Suppose $S \otimes T$ is invertible, so it is injective. Fix $w \in W$ such that $w \neq 0$. If $Sv = 0$ for some $v \in V$, then Exercise 9D9 and Exercise 9D1 give that $(S \otimes T)(v \otimes w) = 0 \otimes Tw = 0$. Therefore, $v \otimes w = 0$ by 3.15. By Exercise 9D1, $v = 0$, so $\ker S = \{0\}$, so S is injective by 3.15. A similar argument shows that T is injective, and thus S and T are invertible by 3.65.

$\Leftarrow$ Suppose S and T are invertible. Then for all $v \in V$ and $w \in W$,

$$(S \otimes T)(S^{-1} \otimes T^{-1})(v \otimes w) = (S \otimes T)(S^{-1}v \otimes T^{-1}w) = SS^{-1}v \otimes TT^{-1}w = v \otimes w.$$

Note that both operators $I_{V\otimes W}$ and $(S\otimes T)(S^{-1}\otimes T^{-1})$ send $v\otimes w$ to itself for all $v\in V$ and $w\in W$, so $(S\otimes T)(S^{-1}\otimes T^{-1}) = I_{V\otimes W}$ by the uniqueness in Exercise 9D9, with I_V and I_W in place of S and T. By 3.68, $(S^{-1}\otimes T^{-1})(S\otimes T) = I_{V\otimes W}$ as well, so $S\otimes T$ is invertible with inverse $S^{-1}\otimes T^{-1}$. □

Solution 9D.11 For all $v_1, v_2 \in V$ and $w_1, w_2 \in W$, we have

$$\begin{aligned}\langle (S\otimes T)(v_1\otimes w_1), v_2\otimes w_2\rangle &= \langle Sv_1\otimes Tw_1, v_2\otimes w_2\rangle\\ &= \langle Sv_1, v_2\rangle\langle Tw_1, w_2\rangle\\ &= \langle v_1, S^*v_2\rangle\langle w_1, T^*w_2\rangle\\ &= \langle v_1\otimes w_1, S^*v_2\otimes T^*w_2\rangle\\ &= \langle v_1\otimes w_1, (S^*\otimes T^*)(v_2\otimes w_2)\rangle,\end{aligned}$$

where the first equality is by Exercise 9D9 applied to S and T; the second and fourth equalities are by 9.80; and the last equality is by Exercise 9D9 applied to S^* and T^*. By 9.74(b), tensors of the form $v\otimes w$ for $v\in V$ and $w\in W$ span $V\otimes W$. By the linearity of $S\otimes T$ and $S^*\otimes T^*$ and by 6.6(a), (d), and (e), the equality $\langle (S\otimes T)x, y\rangle = \langle x, (S^*\otimes T^*)y\rangle$ extends to all $x, y\in V\otimes W$. Therefore, by the uniqueness of the adjoint, $(S\otimes T)^* = S^*\otimes T^*$. □

Solution 9D.12 By 6.35, there exists an orthonormal basis $e_1^k, \dots, e_{n_k}^k$ of V_k for each $k = 1, \dots, m$. For each $J = (j_1, \dots, j_m) \in \{1, \dots, n_1\}\times\cdots\times\{1, \dots, n_m\}$, define $E_J = e_{j_1}^1\otimes\cdots\otimes e_{j_m}^m$. By 9.90, $\mathscr{E} = \{E_J\}_J$ is a basis of $V_1\otimes\cdots\otimes V_m$. By the lemma proved in Exercise 7B24, there exists an inner product on $V_1\otimes\cdots\otimes V_m$ such that $\mathscr{E}$ is an orthonormal basis. Then for any $w\in V_1\otimes\cdots\otimes V_m$, $\langle w, E_J\rangle$ is the coefficient of E_J in the expansion of w in the basis $\mathscr{E}$ by 6.30(a) and 2.28. For any $(v_1, \dots, v_m), (u_1, \dots, u_m) \in V_1\times\cdots\times V_m$, fix $x = v_1\otimes\cdots\otimes v_m$ and $y = u_1\otimes\cdots\otimes u_m$. Define

$$\lambda_J = \prod_{k=1}^m \langle v_k, e_{j_k}^k\rangle \quad\text{and}\quad \mu_J = \prod_{k=1}^m \langle u_k, e_{j_k}^k\rangle.$$

Since each $\phi_k \in V_k'$ in 9.88 is linear, the map $(w_1, \dots, w_m) \mapsto w_1\otimes\cdots\otimes w_m$ is multilinear, so expanding each v_k in x by 6.30(a) gives $x = \sum_J \lambda_J E_J$ and likewise $y = \sum_J \mu_J E_J$. Therefore, by the observation above, 4.4, and 6.30(c),

$$\langle x, y\rangle = \sum_J \lambda_J\overline{\mu_J} = \prod_{k=1}^m\sum_{j=1}^{n_k}\langle v_k, e_j^k\rangle\overline{\langle u_k, e_j^k\rangle} = \prod_{k=1}^m \langle v_k, u_k\rangle,$$

where the last equality follows from 6.30(c) applied in each V_k.

For any inner product that satisfies the condition, applying the condition to the tensors E_J and $E_{J'}$ fixes $\langle E_J, E_{J'}\rangle$ for all J and J'. By 9.90, 6.6(a), (d), and (e), the inner product is determined by its values on pairs $(E_J, E_{J'})$ for all J and J', so the inner product is uniquely determined. □

Solution 9D.13 By 9.90, the list in the statement is a basis of $V_1 \otimes \cdots \otimes V_m$. By Exercise 9D12,

$$\langle e^1_{j_1} \otimes \cdots \otimes e^m_{j_m}, e^1_{\ell_1} \otimes \cdots \otimes e^m_{\ell_m} \rangle = \prod_{k=1}^{m} \langle e^k_{j_k}, e^k_{\ell_k} \rangle = \prod_{k=1}^{m} \delta_{j_k,\ell_k} = \delta_{(j_1,\ldots,j_m),(\ell_1,\ldots,\ell_m)},$$

where $j_k, \ell_k \in \{1, \ldots, n_k\}$ for each $k = 1, \ldots, m$. Therefore, the list is orthonormal, and thus an orthonormal basis. □

Colophon

This edition

This edition was set in LuaLaTeX with `greenbook.sty`, on the memoir class. The text is set in Sabon, the section headings in Jost, the chapter titles in EB Garamond and the chapter numbers in Riesling; the math is set in MathTime Pro 2 Lite. Exercises and solutions are handled by the package xsim, by Clemens Niederberger. The text and the figures (redrawn in TikZ) and the solutions were transcribed from the original PDFs.

Notes on typesetting of the original edition

- This book was typeset in LuaLaTeX by the author, who wrote the LaTeX code to implement the book's design.
- The LaTeX software used for this book was written by Leslie Lamport. The TeX software, which forms the base for LaTeX, was written by Donald Knuth.
- The main text font in this book is the Open Type Format version of TeX Gyre Termes, a font based on Times, which was designed by Stanley Morison and Victor Lardent for the British newspaper *The Times* in 1931.
- The main math font in this book is the Open Type Format version of TeX Gyre Pagella Math, a font based on Palatino, which was designed by Hermann Zapf.
- The sans serif font used for page headings and some other design elements is the Open Type Format version of TeX Gyre Heros, a font based on Helvetica, which was designed by Max Miedinger and Eduard Hoffmann.
- The LuaLaTeX packages fontspec and unicode-math, both written by Will Robertson, were used to manage fonts.
- The LaTeX package fontsize, written by Ivan Valbusa, was used to gracefully change the main fonts to 10.5 point size.
- The figures in the book were produced by *Mathematica*, using *Mathematica* code written by the author. *Mathematica* was created by Stephen Wolfram. The *Mathematica* package MaTeX, written by Szabolcs Horvát, was used to place LaTeX-generated labels in the *Mathematica* figures.
- The LaTeX package graphicx, written by David Carlisle and Sebastian Rahtz, was used to include photos and figures.
- The LaTeX package multicol, written by Frank Mittelbach, was used to get around LaTeX's limitation that two-column format must start on a new page (needed for the Symbol Index and the Index).

- The LaTeX packages Ti*k*Z, written by Till Tantau, and tcolorbox, written by Thomas Sturm, were used to produce the definition boxes and result boxes.
- The LaTeX package color, written by David Carlisle, was used to add appropriate color to various design elements.
- The LaTeX package wrapfig, written by Donald Arseneau, was used to wrap text around the comment boxes.
- The LaTeX package microtype, written by Robert Schlicht, was used to reduce hyphenation and produce more pleasing right justification.