Linear Algebra Done Right

preliminary version of fourth edition
6 February 2023

Sheldon Axler

This document contains the front matter and Chapter 7 of the future fourth edition of *Linear Algebra Done Right*. Suggestions for improvements are most welcome. Please send them to linear@axler.net.

The fourth edition of *Linear Algebra Done Right* will be an Open Access book, which means that the electronic version will be legally free to the world. The print version will be published by Springer and will be reasonably priced. Both the electronic and the print versions will become available around December 2023.

Because this is not the final version of Chapter 7, please do not post this document elsewhere on the web, although it is fine to post a link to it (https://linear.axler.net/).

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

Cover equation: Formula for n^{th} Fibonacci number. Exercise 20 in Section 5D derives this formula by diagonalizing an appropriate operator.

About the Author

Sheldon Axler was valedictorian of his high school in Miami, Florida. He received his AB from Princeton University with highest honors, 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 350 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 a member 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.

He 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.

Contents

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 are in yellow boxes and theorems are in blue boxes (in color versions of the book). Each theorem has a descriptive name.

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

Best wishes for success and enjoyment in learning linear algebra!

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

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

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 audacious title of this book deserves an explanation. Most linear algebra books use determinants to prove that every linear operator on a finite-dimensional complex vector space has an eigenvalue. Determinants are difficult, nonintuitive, and often defined without motivation. To prove the theorem about existence of eigenvalues on complex vector spaces, most books must define determinants, prove that a linear 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 banished to the end of the book, a new route opens to the main goal of linear algebra—understanding the structure of linear operators.

This book starts at the beginning of the subject, with no prerequisites other than the usual demand for suitable mathematical maturity. A few examples and exercises involve calculus concepts such as continuity, differentiation, or 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: Linear maps are introduced in this chapter. The key result here is the fundamental theorem of linear maps (3.21): if T is a linear map on V, then $\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$. Quotient spaces and duality are topics in this chapter at a higher level of abstraction than other parts of the book; these topics can be skipped without running into problems elsewhere in the book.

- Chapter 4: The part of the theory of polynomials that will be needed to understand linear operators is presented in this chapter. This chapter contains no linear algebra. It can be covered quickly, especially if your students are already familiar with these results.
- Chapter 5: The idea of studying a linear operator by restricting it to small subspaces leads to eigenvectors in the early part of this chapter. The highlight of this chapter is a simple proof that on complex vector spaces, eigenvalues always exist. This result is then used to show that each linear operator on a complex vector space has an upper-triangular matrix with respect to some basis. 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.
- 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.
- 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, the QR factorization, the singular value decomposition, 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. The chapter concludes with a proof that every linear operator on a complex vector space can be put into Jordan form.
- Chapter 9: Operators on real vector spaces occupy center stage in this chapter. Here the main technique is complexification, which is a natural extension of an operator on a real vector space to an operator on a complex vector space. Complexification allows results about complex vector spaces to be transferred easily to real vector spaces.
- Chapter 10: The trace and determinant (on complex vector spaces) are defined in this chapter as the sum of the eigenvalues and the product of the eigenvalues, with each eigenvalue included in the sum or product as many times as its multiplicity. These easy-to-remember definitions are possible because we have already proved that sufficient eigenvalues exist without using determinants Some standard theorems about determinants now become much clearer.

This book usually develops linear algebra simultaneously for real and complex vector spaces by letting **F** denote either the real or the complex numbers. If you and your students prefer to think of **F** as an arbitrary field, then see the comments at the end of Section 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 the first eight chapters is a good goal for a one-semester course. If you must reach Chapter 10, then consider covering Chapters 4 and 9 quickly in a half hour each, as well as skipping the material on quotient spaces and duality in Chapter 3.

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

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 ten crucial results will have a solid foundation in linear algebra.

- any two bases of V have same length (2.34)
- fundamental theorem of linear maps (3.21)
- existence of eigenvalues if $\mathbf{F} = \mathbf{C}$ (5.19)
- upper-triangular form always exists if $\mathbf{F} = \mathbf{C}$ (5.47)
- Cauchy–Schwarz inequality (6.14)
- Parseval's identity for orthonormal bases (6.30)
- Gram–Schmidt procedure (6.32)
- spectral theorem (7.29 and 7.31)
- singular value decomposition (7.67)
- generalized eigenvector decomposition theorem when $\mathbf{F} = \mathbf{C}$ (8.22)

Major improvements and additions for the fourth edition

- 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 subsection on QR 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 formatting to improve the appearance of the book. For example, the definition and result boxes now have rounded corners instead of right-angle corners, for a gentler appearance. The main font size has been reduced from 11 point to 10.5 point.

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

Best wishes for teaching a successful linear algebra class!

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

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

Acknowledgments

I owe a huge intellectual debt to the many mathematicians who created linear algebra over the past two centuries. The results in this book belong to the common heritage of mathematics. A special case of a theorem may first have been proved long ago, then sharpened and improved by many mathematicians in different time periods. Bestowing proper credit on all the contributors would be a difficult task that I have not undertaken. In no case should the reader assume that any theorem presented here represents my original contribution. However, in writing this book I tried to think about the best way to present linear algebra and to prove its theorems.

MORE TO COME

Sheldon Axler

Chapter 7

Operators on Inner Product Spaces

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 products 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 lead to unitary matrices and the QR factorization. The spectral theorem will lead to the popular singular value decomposition, which leads to the definition of the norm of a linear map and to the polar decomposition.

A new assumption for this chapter is listed in the second bullet point below.

standing assumptions for this chapter

- **F** denotes **R** or **C**.
- V and W are nonzero finite-dimensional inner product spaces over **F**.

Market square in Lviv, a city that has had several names and has been in several countries because of changing international borders. From 1772 until 1918, the city was in Austria and was called Lemberg. Between World War I and World War II, the city was in Poland and was called Lwów. During this time, mathematicians in Lwów, particularly Stefan Banach (1892–1945) and his colleagues, developed the basic results of modern functional analysis, using tools of analysis to study infinite-dimensional vector spaces. Since the end of World War II, Lviv has been in Ukraine, which was part of the Soviet Union until Ukraine became an independent country in 1991.

7A Self-Adjoint and Normal Operators

Adjoints

7.1 definition: *adjoint, T^**

Suppose $T \in \mathcal{L}(V, W)$. The *adjoint* of T is the function $T^* \colon W \to V$ such that

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for every $v \in V$ and every $w \in W$.

*The word **adjoint** has another meaning in linear algebra. We do not need the second meaning in this book. In case you encounter the second meaning for adjoint elsewhere, be warned that the two meanings for adjoint are unrelated to each other.*

To see why the definition above makes sense, suppose $T \in \mathcal{L}(V, W)$. Fix $w \in W$. Consider the linear functional

$$v \mapsto \langle Tv, w \rangle$$

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 (6.43), there exists a unique vector in V such that this linear functional is given by taking the inner product with it. We call this unique vector $T^* w$. In other words, $T^* w$ is the unique vector in V such that $\langle Tv, w \rangle = \langle v, T^* w \rangle$ for every $v \in V$.

7.2 example: *adjoint of a linear map from $\mathbf{R}^3$ to $\mathbf{R}^2$*

Define $T \colon \mathbf{R}^3 \to \mathbf{R}^2$ by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

To compute T^*, fix a point $(y_1, y_2) \in \mathbf{R}^2$. Then for every $(x_1, x_2, x_3) \in \mathbf{R}^3$ we have

$$\begin{aligned} \langle (x_1, x_2, x_3), T^*(y_1, y_2) \rangle &= \langle T(x_1, x_2, x_3), (y_1, y_2) \rangle \\ &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2) \rangle \\ &= x_2 y_1 + 3x_3 y_1 + 2x_1 y_2 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1) \rangle. \end{aligned}$$

Thus

$$\langle (x_1, x_2, x_3), T^*(y_1, y_2) - (2y_2, y_1, 3y_1) \rangle = 0$$

for all $(x_1, x_2, x_3) \in \mathbf{R}^3$. Taking $(x_1, x_2, x_3) = T^*(y_1, y_2) - (2y_2, y_1, 3y_1)$ and using the property that $(0, 0, 0)$ is the only element of $\mathbf{R}^3$ whose inner product with itself equals 0, we have

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

7.3 example: *adjoint of a linear map with range of dimension at most* 1

Fix $u \in V$ and $x \in W$. Define $T \in \mathcal{L}(V, W)$ by

$$Tv = \langle v, u\rangle x$$

for each $v \in V$. To compute T^*, fix $w \in W$. Then for every $v \in V$ we have

$$\begin{aligned}\langle v, T^*w\rangle &= \langle Tv, w\rangle \\ &= \big\langle \langle v, u\rangle x, w\big\rangle \\ &= \langle v, u\rangle\langle x, w\rangle \\ &= \big\langle v, \langle w, x\rangle u\big\rangle.\end{aligned}$$

Thus

$$\big\langle v, T^*w - \langle w, x\rangle u\big\rangle = 0$$

for all $v \in V$. Taking $v = T^*w - \langle w, x\rangle u$ and using the property that 0 is the only element of V whose inner product with itself equals 0, we have

$$T^*w = \langle w, x\rangle u.$$

In the two examples above, T^* turned out to be not just a function but a linear map. This is true in general, as shown by the next result.

The two examples above and the proofs of the next two results use a common technique for computing T^*: start with $\langle v, T^*w\rangle$, flip T^* from the second slot to become T in the first slot, then manipulate to get just v in the first slot.

7.4 *adjoint of a linear map is a linear map*

If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.

Proof Suppose $T \in \mathcal{L}(V, W)$. Fix $w_1, w_2 \in W$. If $v \in V$, then

$$\begin{aligned}\big\langle v, T^*(w_1 + w_2)\big\rangle &= \langle Tv, w_1 + w_2\rangle \\ &= \langle Tv, w_1\rangle + \langle Tv, w_2\rangle \\ &= \big\langle v, T^*w_1\big\rangle + \big\langle v, T^*w_2\big\rangle \\ &= \big\langle v, T^*w_1 + T^*w_2\big\rangle,\end{aligned}$$

which shows that $T^*(w_1 + w_2) = T^*w_1 + T^*w_2$.

Fix $w \in W$ and $\lambda \in \mathbf{F}$. If $v \in V$, then

$$\begin{aligned}\big\langle v, T^*(\lambda w)\big\rangle &= \langle Tv, \lambda w\rangle \\ &= \overline{\lambda}\langle Tv, w\rangle \\ &= \overline{\lambda}\big\langle v, T^*w\big\rangle \\ &= \big\langle v, \lambda T^*w\big\rangle,\end{aligned}$$

which shows that $T^*(\lambda w) = \lambda T^*w$.

Thus T^* is a linear map, as desired. ∎

7.5 *properties of the adjoint*

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

(a) $(S+T)^* = S^* + T^*$ for all $S \in \mathcal{L}(V, W)$;

(b) $(\lambda T)^* = \overline{\lambda} T^*$ for all $\lambda \in \mathbf{F}$;

(c) $(T^*)^* = T$;

(d) $(ST)^* = T^* S^*$ for all $S \in \mathcal{L}(W, U)$ (here U is an inner product space over $\mathbf{F}$);

(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

(a) Suppose $S \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\begin{aligned}\langle v, (S+T)^* w\rangle &= \langle (S+T)v, w\rangle \\ &= \langle Sv, w\rangle + \langle Tv, w\rangle \\ &= \langle v, S^* w\rangle + \langle v, T^* w\rangle \\ &= \langle v, S^* w + T^* w\rangle.\end{aligned}$$

Thus $(S+T)^* w = S^* w + T^* w$, as desired.

(b) Suppose $\lambda \in \mathbf{F}$. If $v \in V$ and $w \in W$, then

$$\langle v, (\lambda T)^* w\rangle = \langle \lambda T v, w\rangle = \lambda \langle Tv, w\rangle = \lambda \langle v, T^* w\rangle = \langle v, \overline{\lambda} T^* w\rangle.$$

Thus $(\lambda T)^* w = \overline{\lambda} T^* w$, as desired.

(c) If $v \in V$ and $w \in W$, then

$$\langle w, (T^*)^* v\rangle = \langle T^* w, v\rangle = \overline{\langle v, T^* w\rangle} = \overline{\langle Tv, w\rangle} = \langle w, Tv\rangle.$$

Thus $(T^*)^* v = Tv$, as desired.

(d) Suppose $S \in \mathcal{L}(W, U)$. If $v \in V$ and $u \in U$, then

$$\langle v, (ST)^* u\rangle = \langle STv, u\rangle = \langle Tv, S^* u\rangle = \langle v, T^*(S^* u)\rangle.$$

Thus $(ST)^* u = T^*(S^* u)$, as desired.

(e) If $v, u \in V$, then

$$\langle v, I^* u\rangle = \langle Iv, u\rangle = \langle v, u\rangle.$$

Thus $I^* u = u$, as desired.

(f) Suppose T is invertible. Take adjoints of both sides of the equation $T^{-1}T = I$ and use parts (d), (e) to show $T^*(T^{-1})^* = I$. Similarly, $(T^{-1})^* T^* = I$. ■

Parts (a) and (b) of the result above show that if $\mathbf{F} = \mathbf{R}$ then the map $T \mapsto T^*$ is a linear map of $\mathcal{L}(V)$ to itself. However, if $\mathbf{F} = \mathbf{C}$ and $\dim V > 0$, then this map is not linear because of the complex conjugate that appears in part (b).

The next result shows the relationship between the null space and the range of a linear map and its adjoint.

7.6 *null space and range of T^**

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

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Proof We begin by proving (a). Let $w \in W$. Then

$$\begin{aligned} w \in \operatorname{null} T^* &\iff T^*w = 0 \\ &\iff \langle v, T^*w \rangle = 0 \text{ for all } v \in V \\ &\iff \langle Tv, w \rangle = 0 \text{ for all } v \in V \\ &\iff w \in (\operatorname{range} T)^\perp. \end{aligned}$$

Thus $\operatorname{null} T^* = (\operatorname{range} T)^\perp$, proving (a).

If we take the orthogonal complement of both sides of (a), we get (d), where we have used 6.54. 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.

7.7 definition: *conjugate transpose, A^**

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, $(A^*)_{j,k} = \overline{A_{k,j}}$.

7.8 example: *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

$$\begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}.$$

If a matrix has only real entries, then its conjugate transpose equals its ***transpose****, which is the matrix obtained by interchanging rows and columns.*

The next result shows how to compute the matrix of T^* from the matrix of T.

Caution: Remember that the result below applies only when we are dealing with orthonormal bases. With respect to nonorthonormal bases, the matrix of T^* does not necessarily equal the conjugate transpose of the matrix of T.

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.

7.9 *matrix of T^* equals conjugate transpose of matrix of T*

Let $T \in \mathcal{L}(V, W)$. 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 $\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$ is the conjugate transpose of $\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$. In other words,

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^*.$$

Proof In this proof, we will write $\mathcal{M}(T)$ and $\mathcal{M}(T^*)$ instead of the longer expressions $\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$ and $\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$.

Recall that we obtain the k^{th} column of $\mathcal{M}(T)$ by writing Te_k as a linear combination of the f_j's; the scalars used in this linear combination then become the k^{th} column of $\mathcal{M}(T)$. Because $f_1, \dots, f_m$ is an orthonormal basis of W, we know how to write Te_k as a linear combination of the f_j's [see 6.30(a)]:

$$Te_k = \langle Te_k, f_1 \rangle f_1 + \cdots + \langle Te_k, f_m \rangle f_m.$$

Thus

the entry in row j, column k, of $\mathcal{M}(T)$ is $\langle Te_k, f_j \rangle$.

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 $\mathcal{M}(T^*)$ is $\langle T^* f_k, e_j \rangle$, which equals $\langle f_k, Te_j \rangle$, which equals $\overline{\langle Te_j, f_k \rangle}$, which equals the complex conjugate of the entry in row k, column j, of $\mathcal{M}(T)$. Thus $\mathcal{M}(T^*) = (\mathcal{M}(T))^*$. ■

The Riesz representation theorem as stated in 6.60 provides an identification of V with its dual space V' defined in 3.90. Under this identification, the orthogonal complement $U^\perp$ of a subset $U \subset V$ 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 3.104 and 6.53.

Suppose $T\colon V \to W$ is a linear map. Under the identification of V with V' and the identification of W with W', the adjoint map $T^*\colon W \to V$ corresponds to the dual map $T'\colon W' \to V'$ defined in 3.90, as Exercise 11 asks you to verify. Under this identification, the formulas for null T^* and range T^* [7.6(a) and (b)] then become identical with the formulas for null T' and range T' [3.107(a) and 3.109(b)]. Furthermore, the theorem about the matrix of T^* (7.9) is analogous to the theorem about the matrix of T' (3.114).

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. Thus instead of considering linear maps from V to W, we will be focusing on linear maps from V to V; recall that such linear maps are called operators.

7.10 definition: *self-adjoint*

An operator $T \in \mathcal{L}(V)$ is called *self-adjoint* if $T = T^*$. In other words, $T \in \mathcal{L}(V)$ is self-adjoint if and only if

$$\langle Tv, w \rangle = \langle v, Tw \rangle$$

for all $v, w \in V$.

7.11 example: *determining whether T is self-adjoint from its matrix*

Suppose $c \in \mathbf{F}$ and T is the operator on $\mathbf{F}^2$ whose matrix (with respect to the standard basis) is

$$\mathcal{M}(T) = \begin{pmatrix} 2 & c \\ 3 & 7 \end{pmatrix}.$$

The matrix of T^* (with respect to the standard basis) is

$$\mathcal{M}(T^*) = \begin{pmatrix} 2 & 3 \\ \bar{c} & 7 \end{pmatrix}.$$

Thus $\mathcal{M}(T) = \mathcal{M}(T^*)$ if and only if $c = 3$. In other words, the operator T is self-adjoint if and only if $c = 3$.

You should verify that the sum of two self-adjoint operators is self-adjoint and that the product of a real scalar and a self-adjoint operator is self-adjoint.

A good analogy to keep in mind is that the adjoint on $\mathcal{L}(V)$ plays a role similar to the complex conjugate on $\mathbf{C}$. A complex number z is real if and only if $z = \bar{z}$; thus a self-adjoint operator ($T = T^*$) is analogous to a real number.

We will see that the analogy discussed above is reflected in some important properties of self-adjoint operators, beginning with eigenvalues in the next result.

If $\mathbf{F} = \mathbf{R}$, then by definition every eigenvalue is real, so the next result is interesting only when $\mathbf{F} = \mathbf{C}$.

7.12 *eigenvalues of self-adjoint operators are real*

Every eigenvalue of a self-adjoint operator is real.

Proof Suppose T is a self-adjoint operator on V. Let λ be an eigenvalue of T, and let v be a nonzero vector in V such that $Tv = \lambda v$. Then

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \langle v, \lambda v \rangle = \bar{\lambda} \|v\|^2.$$

Thus $\lambda = \bar{\lambda}$, which means that λ is real, as desired. ■

The next result is false for real inner product spaces. As an example, consider the operator $T \in \mathcal{L}(\mathbf{R}^2)$ that is a counterclockwise rotation of 90° around the origin; thus $T(x, y) = (-y, x)$. Obviously Tv is orthogonal to v for every $v \in \mathbf{R}^2$, even though $T \neq 0$.

7.13 *Tv is orthogonal to v for all v $\iff$ $T = 0$ (assuming $\mathbf{F} = \mathbf{C}$)*

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. 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 $u \in U$ (take $w = Tu$). Hence $T = 0$, as desired. ■

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.

The next result provides another good example of how self-adjoint operators behave like real numbers.

7.14 *$\langle Tv, v\rangle$ is real for all v $\iff$ T is self-adjoint (assuming $\mathbf{F} = \mathbf{C}$)*

Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then

$$T \text{ is self-adjoint} \iff \langle Tv, v\rangle \in \mathbf{R} \text{ for every } v \in V.$$

Proof If $v \in V$, then

7.15 $$\langle T^*v, v\rangle = \langle v, Tv\rangle = \overline{\langle Tv, v\rangle}.$$

Now

$$\begin{aligned} T \text{ is self-adjoint} &\iff T - T^* = 0 \\ &\iff \langle (T - T^*)v, v\rangle = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle \in \mathbf{R} \text{ for every } v \in V, \end{aligned}$$

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.

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

7.17 $$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4},$$

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

7.18 definition: *normal*

- An operator on an inner product space is called *normal* if it commutes with its adjoint.
- In other words, $T \in \mathcal{L}(V)$ is normal if $TT^* = T^*T$.

Obviously every self-adjoint operator is normal, because if T is self-adjoint then $T^* = T$.

7.19 example: *an operator that is not self-adjoint but is normal*

Let T be the operator on $\mathbf{F}^2$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

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

$$\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}, \text{ which equals } \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Similarly, the matrix of T^*T equals

$$\begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}, \text{ which equals } \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix}.$$

Because TT^* and T^*T have the same matrix, we see that $TT^* = T^*T$. Thus T is normal.

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.

7.20 *T is normal if and only if $\|Tv\| = \|T^*v\|$ for all v*

Suppose $T \in \mathcal{L}(V)$. Then

$$T \text{ is normal } \iff \|Tv\| = \|T^*v\| \text{ for every } v \in V.$$

Proof We have

$$\begin{aligned} T \text{ is normal } &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0 \text{ for every } v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \text{ for every } v \in V \\ &\iff \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle \text{ for every } v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \text{ for every } v \in V \\ &\iff \|Tv\| = \|T^*v\| \text{ for every } v \in V, \end{aligned}$$

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 part (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, part (e) of the next result implies that a normal operator and its adjoint have the same eigenvectors.

7.21 *range, null space, and eigenvectors of a normal operator*

Suppose $T \in \mathcal{L}(V)$ is normal. Then

(a) $\operatorname{null} T = \operatorname{null} T^*$;

(b) $\operatorname{range} T = \operatorname{range} T^*$;

(c) $V = \operatorname{null} T \oplus \operatorname{range} T$;

(d) $T - \lambda I$ is normal for every $\lambda \in \mathbf{F}$;

(e) $Tv = \lambda v$ if and only if $T^* v = \overline{\lambda} v$ (for each $v \in V$ and each $\lambda \in \mathbf{F}$).

Proof

(a) Suppose $v \in V$. Then

$$v \in \operatorname{null} T \iff \|Tv\| = 0 \iff \|T^* v\| = 0 \iff v \in \operatorname{null} T^*,$$

where the middle equivalence above follows from 7.20. Thus $\operatorname{null} T = \operatorname{null} T^*$.

(b) We have

$$\operatorname{range} T = (\operatorname{null} T^*)^\perp = (\operatorname{null} T)^\perp = \operatorname{range} T^*,$$

where the first equality comes from 7.6(d), the second equality comes from part (a) of this result, and the third equality comes from 7.6(b).

(c) We have

$$V = (\operatorname{null} T) \oplus (\operatorname{null} T)^\perp = \operatorname{null} T \oplus \operatorname{range} T^* = \operatorname{null} T \oplus \operatorname{range} T,$$

where the first equality comes from 6.50, the second equality comes from 7.6(b), and the third equality comes from part (b) of this result.

(d) Suppose $\lambda \in \mathbf{F}$. 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, and hence $T - \lambda I$ is normal.

(e) Suppose $v \in V$ and $\lambda \in \mathbf{F}$. Then part (d) of this result 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.

7.22 *orthogonal eigenvectors for normal operators*

Suppose $T \in \mathcal{L}(V)$ is normal. Then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.

Proof Suppose α, β are distinct eigenvalues of T, with corresponding eigenvectors u, v. Thus $Tu = \alpha u$ and $Tv = \beta v$. From 7.21(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 $\mathbf{F} = \mathbf{C}$. However, see Exercise 27 for a version that makes sense when $\mathbf{F} = \mathbf{C}$ and when $\mathbf{F} = \mathbf{R}$.

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Under the analogy between $\mathcal{L}(V)$ and $\mathbf{C}$, with the adjoint on $\mathcal{L}(V)$ playing a similar role to the complex conjugate on $\mathbf{C}$, the operators Q and S as defined by 7.24 correspond to the real and imaginary parts of T. Thus the informal title given below to 7.23 should make sense.

7.23 *T is normal $\iff$ the real and imaginary parts of T commute*

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then T is normal if and only if there exist commuting self-adjoint operators Q and S such that $T = Q + iS$.

Proof First suppose T is normal. Let

7.24
$$Q = \frac{T + T^*}{2} \quad \text{and} \quad S = \frac{T - T^*}{2i}.$$

Then Q and S are self-adjoint and $T = Q + iS$. A quick computation shows that

7.25
$$QS - SQ = \frac{T^*T - TT^*}{2i}.$$

Because T is normal, the right side of the equation above equals 0. Thus the operators Q and S commute, as desired.

To prove the implication in the other direction, now suppose there exist commuting self-adjoint operators Q and S such that $T = Q + iS$. Then $T^* = Q - iS$. Adding the last two equations and then dividing by 2 produces the equation for Q in 7.24. Subtracting the last two equations and then dividing by $2i$ produces the equation for S in 7.24. Now 7.24 implies 7.25. Because S and Q commute, 7.25 implies that T is normal, as desired. ■

Exercises 7A

1 Suppose n is a positive integer. Define $T \in \mathcal{L}(\mathbf{F}^n)$ by

$$T(z_1, \dots, z_n) = (0, z_1, \dots, z_{n-1}).$$

Find a formula for $T^*(z_1, \dots, z_n)$.

2 Suppose $T \in \mathcal{L}(V, W)$. Prove that

$$T = 0 \iff T^* = 0 \iff T^*T = 0 \iff TT^* = 0.$$

3 Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Prove that

$$\lambda \text{ is an eigenvalue of } T \iff \overline{\lambda} \text{ is an eigenvalue of } T^*.$$

4 Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V. Prove that

$$U \text{ is invariant under } T \iff U^\perp \text{ is invariant under } T^*.$$

5 Suppose $T \in \mathcal{L}(V, W)$. 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.

6 Suppose $T \in \mathcal{L}(V)$ and $\|T^*v\| \le \|Tv\|$ 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.

7 Suppose $T \in \mathcal{L}(V, W)$. Prove that

(a) T is injective $\iff$ T^* is surjective;
(b) T is surjective $\iff$ T^* is injective.

8 Prove that if $T \in \mathcal{L}(V, W)$, then

(a) $\dim \operatorname{null} T^* = \dim \operatorname{null} T + \dim W - \dim V$;
(b) $\dim \operatorname{range} T^* = \dim \operatorname{range} T$.

9 Suppose A is an m-by-n matrix with entries in $\mathbf{F}$. Use part (b) of Exercise 8 to prove that the row rank of A equals the column rank of A, using adjoints instead of the duality that was used in 3.118 (see 3.115 for definitions).

10 Prove that the product of two self-adjoint operators on V is self-adjoint if and only if the two operators commute.

11 Suppose $T\colon V \to W$ is a linear map. Show that under the standard identification of V with V' (see 6.60) and the corresponding identification of W with W', the adjoint map $T^*\colon W \to V$ corresponds to the dual map $T'\colon W' \to V'$. More precisely, show that $T'(\varphi_w) = \varphi_{T^*w}$ for all $w \in W$, where φ is as in 6.60.

12 Define an inner product on $\mathcal{P}_2(\mathbf{R})$ by $\langle p, q\rangle = \int_0^1 pq$. Define an operator $T \in \mathcal{L}(\mathcal{P}_2(\mathbf{R}))$ by $T(a_0 + a_1x + a_2x^2) = a_1x$.

(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.

13 Suppose $T \in \mathcal{L}(V)$ is invertible. Prove that

(a) T is self-adjoint $\iff$ T^{-1} is self-adjoint;

(b) T is normal $\iff$ T^{-1} is normal.

14 Suppose V is a real inner product space.

(a) Show that the set of self-adjoint operators on V is a subspace of $\mathcal{L}(V)$.

(b) What is the dimension of the subspace of $\mathcal{L}(V)$ in part (a) [in terms of $\dim V$]?

15 Suppose V is a complex inner product space with $V \neq \{0\}$. Show that the set of self-adjoint operators on V is not a subspace of $\mathcal{L}(V)$.

16 Suppose $\dim V \geq 2$. Show that the set of normal operators on V is not a subspace of $\mathcal{L}(V)$.

17 Suppose $P \in \mathcal{L}(V)$ is such that $P^2 = P$. Prove that there is a subspace U of V such that $P = P_U$ if and only if P is self-adjoint.

18 Suppose $D\colon \mathcal{P}_8(\mathbf{R}) \to \mathcal{P}_8(\mathbf{R})$ is the differentiation operator defined by $Dp = p'$. Prove that there does not exist an inner product on $\mathcal{P}_8(\mathbf{R})$ that makes D a normal operator.

19 Give an example of an operator $T \in \mathcal{L}(\mathbf{R}^3)$ such that T is normal but not self-adjoint.

20 Suppose T is a normal operator on V. Suppose also that $v, w \in V$ satisfy the equations

$$\|v\| = \|w\| = 2, \quad Tv = 3v, \quad Tw = 4w.$$

Show that $\|T(v + w)\| = 10$.

21 Suppose $T \in \mathcal{L}(V)$ 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 $\mathbf{F} = \mathbf{R}$.

22 Fix $u, x \in V$. Define $T \in \mathcal{L}(V)$ by $Tv = \langle v, u \rangle x$ for every $v \in V$.

(a) Show that $T^* w = \langle w, x \rangle u$ for every $w \in V$.
(b) 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.
(c) Prove that T is normal if and only if the list u, x is linearly dependent.

23 Suppose $T \in \mathcal{L}(V)$ is normal. Prove that

$$\operatorname{null} T^k = \operatorname{null} T \quad \text{and} \quad \operatorname{range} T^k = \operatorname{range} T$$

for every positive integer k.

24 Suppose V is an inner product space and $T \in \mathcal{L}(V)$ is normal. Prove that if $\lambda \in \mathbf{F}$, then the minimal polynomial of T is not a polynomial multiple of $(x - \lambda)^2$.

25 Prove or give counterexample: If $T \in \mathcal{L}(V)$ and there exists 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.

26 Suppose that $T \in \mathcal{L}(\mathbf{F}^3)$ is normal and $T(1, 1, 1) = (2, 2, 2)$. Suppose $(z_1, z_2, z_3) \in \operatorname{null} T$. Prove that $z_1 + z_2 + z_3 = 0$.

27 An operator $S \in \mathcal{L}(V)$ is called *skew* if $S^* = -S$. Suppose $T \in \mathcal{L}(V)$. Prove that T is normal if and only if there exist commuting operators Q and S such that Q is self-adjoint, S is a skew operator, and $T = Q + S$.

28 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

$$V = \operatorname{span}(1, \cos x, \cos 2x, \dots, \cos nx, \sin x, \sin 2x, \dots, \sin nx).$$

(a) Define $D \in \mathcal{L}(V)$ by $Df = f'$. Show that $D^* = -D$. Conclude that D is normal but not self-adjoint.
(b) Define $T \in \mathcal{L}(V)$ by $Tf = f''$. Show that T is self-adjoint.

29 Suppose $\mathbf{F} = \mathbf{R}$. Find all eigenvalues of the operator $\mathcal{A} \in \mathcal{L}(\mathcal{L}(V))$ defined by $\mathcal{A}T = T^*$ for each $T \in \mathcal{L}(V)$.

30 Suppose $T \in \mathcal{L}(V)$. Prove that

(a) T is self-adjoint $\iff$ $T^\dagger$ is self-adjoint;
(b) T is normal $\iff$ $T^\dagger$ is normal.

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 $T \in \mathcal{L}(V)$ such that there is an orthonormal basis of V consisting of eigenvectors of T. Our goal in this section is to prove the spectral theorem, which characterizes these operators as the self-adjoint operators when $\mathbf{F} = \mathbf{R}$ and as the normal operators when $\mathbf{F} = \mathbf{C}$.

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 $\mathbf{F}$, 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 $b, c \in \mathbf{R}$ 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.

7.26 *invertible quadratic expressions*

Suppose $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbf{R}$ are such that $b^2 < 4c$. Then

$$T^2 + bT + cI$$

is an invertible operator.

Proof Let v be a nonzero vector in V. Then

$$\begin{aligned}\langle (T^2 + bT + cI)v, v\rangle &= \langle T^2 v, v\rangle + b\langle Tv, v\rangle + c\langle v, v\rangle \\ &= \langle Tv, Tv\rangle + b\langle Tv, v\rangle + c\|v\|^2 \\ &\geq \|Tv\|^2 - |b|\,\|Tv\|\,\|v\| + c\|v\|^2 \\ &= \left(\|Tv\| - \frac{|b|\,\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\ &> 0,\end{aligned}$$

where the third line above holds by the Cauchy–Schwarz inequality (6.14). 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 3.54). ■

The next result will be a key tool in our proof of the real spectral theorem.

7.27 *minimal polynomial of self-adjoint operator*

Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Then the minimal polynomial of T equals $(z - \lambda_1)\cdots(z - \lambda_m)$ for some $\lambda_1, \ldots, \lambda_m \in \mathbf{R}$.

Proof First suppose $\mathbf{F} = \mathbf{C}$. The zeros of the minimal polynomial of T are the eigenvalues of T [by 5.27(a)]. All the 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 $\mathbf{F} = \mathbf{R}$. By the factorization of a polynomial over $\mathbf{R}$ (see 4.16) there exist $\lambda_1, \ldots, \lambda_m \in \mathbf{R}$ and $b_1, \ldots, b_N, c_1, \ldots, c_N \in \mathbf{R}$ with $b_k^2 < 4c_k$ for each k such that the minimal polynomial of T equals

7.28 $$(z - \lambda_1)\cdots(z - \lambda_m)(z^2 + b_1 z + c_1)\cdots(z^2 + b_N z + c_N),$$

here either m or N might equal 0, meaning that there are no terms of the corresponding form. Because the polynomial above is the minimal polynomial of T, we have

$$(T - \lambda_1 I)\cdots(T - \lambda_m I)(T^2 + b_1 T + c_1 I)\cdots(T^2 + b_N T + c_N I) = 0.$$

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, violating the minimality of the degree of the polynomial in 7.28 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 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.

7.29 *real spectral theorem*

Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent.

(a) T is self-adjoint

(b) 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 6.38 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. In other words, 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]. ■

7.30 example: *an orthonormal basis of eigenvectors for an operator*

Consider the self-adjoint operator T on $\mathbf{R}^3$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 14 & -13 & 8 \\ -13 & 14 & 8 \\ 8 & 8 & -7 \end{pmatrix}.$$

As you can verify,

$$\frac{(1,-1,0)}{\sqrt{2}}, \frac{(1,1,1)}{\sqrt{3}}, \frac{(1,1,-2)}{\sqrt{6}}$$

is an orthonormal basis of $\mathbf{R}^3$ consisting of eigenvectors of T, and with respect to this basis, the matrix of T is the diagonal matrix

$$\begin{pmatrix} 27 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & -15 \end{pmatrix}.$$

See Exercise 15 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.

7.31 *complex spectral theorem*

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. 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 (6.39), 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

7.32
$$\mathcal{M}(T, (e_1, \dots, e_n)) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}.$$

We will show that this matrix is actually a diagonal matrix.

We see from the matrix above that

$$\|Te_1\|^2 = |a_{1,1}|^2$$
$$\|T^*e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2.$$

Because T is normal, $\|Te_1\| = \|T^*e_1\|$ (see 7.20). Thus the two equations above imply that all entries in the first row of the matrix in 7.32, 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 the 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 Exercise 8 for an alternative proof that (a) implies (b) in the result above.

See Exercise 11 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 next example.

7.33 example: *an orthonormal basis of eigenvectors for an operator*

Consider the operator $T \in \mathcal{L}(\mathbf{C}^2)$ 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,

$$\frac{1}{\sqrt{2}}(i, 1), \frac{1}{\sqrt{2}}(-i, 1)$$

is an orthonormal basis of $\mathbf{C}^2$ consisting of eigenvectors of T, and with respect to this basis the matrix of T is the diagonal matrix

$$\begin{pmatrix} 2 + 3i & 0 \\ 0 & 2 - 3i \end{pmatrix}.$$

Exercises 7B

1 Prove or disprove: There exists a diagonalizable operator $T \in \mathcal{L}(\mathbf{C}^3)$ such that T is not normal (with respect to the usual inner product).

2 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.

3 Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$ is a normal operator such that $T^9 = T^8$. Prove that T is self-adjoint and $T^2 = T$.

4 Give an example of an operator T on a complex vector space such that $T^9 = T^8$ but $T^2 \neq T$.

5 Suppose V is a complex inner product space. Prove that every normal operator on V has a square root.

*An operator $S \in \mathcal{L}(V)$ is called a **square root** of $T \in \mathcal{L}(V)$ if $S^2 = T$.*

6 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is normal if and only if there exists a polynomial $p \in \mathcal{P}(\mathbf{C})$ such that $T^* = p(T)$.

7 Suppose V is a complex vector space, $T \in \mathcal{L}(V)$ is normal, and $S \in \mathcal{L}(V)$ is such that $ST = TS$. Prove that $ST^* = T^*S$.

The result in this exercise is called Fuglede's theorem.

8 Without using the complex spectral theorem, use the version of Schur's theorem that applies to two commuting operators (take $\mathcal{E} = \{T, T^*\}$ in Exercise 19 in Section 6B) to give a different proof that if $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal, then T has a diagonal matrix with respect to some orthonormal basis of V.

9 Suppose $\mathbf{F} = \mathbf{R}$ and $T \in \mathcal{L}(V)$. Prove that T is self-adjoint if and only if all pairs of eigenvectors corresponding to distinct eigenvalues of T are orthogonal and

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

10 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Prove that T is normal if and only if all pairs of eigenvectors corresponding to distinct eigenvalues of T are orthogonal and

$$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

11 Suppose $\mathbf{F} = \mathbf{C}$ and $\mathcal{E} \subset \mathcal{L}(V)$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix if and only if S and T are commuting normal operators for all $S, T \in \mathcal{E}$.

This exercise extends the complex spectral theorem to the context of a collection of commuting normal operators.

12 Suppose $T \in \mathcal{L}(V)$ 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 $T|_U \in \mathcal{L}(U)$ is self-adjoint.
(c) Prove that $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.

13 Suppose $T \in \mathcal{L}(V)$ 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 $T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are normal operators.

14 Suppose that T is a self-adjoint operator on a finite-dimensional inner product space and that 2 and 3 are the only eigenvalues of T. Prove that $T^2 - 5T + 6I = 0$.

15 Suppose $\mathbf{F} = \mathbf{R}$ and $\mathcal{E} \subset \mathcal{L}(V)$. Prove that there is an orthonormal basis of V with respect to which every element of $\mathcal{E}$ has a diagonal matrix if and only if S and T are commuting self-adjoint operators for all $S, T \in \mathcal{E}$.

This exercise extends the real spectral theorem to the context of a collection of commuting self-adjoint operators.

16 Give an example of a real inner product space V and $T \in \mathcal{L}(V)$ and real numbers b, c with $b^2 < 4c$ such that $T^2 + bT + cI$ is not invertible.

This exercise shows that the hypothesis that T is self-adjoint cannot be deleted in 7.26, even for real vector spaces.

17 Give an example of an operator $T \in \mathcal{L}(\mathbf{C}^3)$ such that 2 and 3 are the only eigenvalues of T and $T^2 - 5T + 6I \neq 0$.

18 Prove or give counterexample: Every self-adjoint operator on V has a cube root.

An operator $S \in \mathcal{L}(V)$ is called a ***cube root*** *of $T \in \mathcal{L}(V)$ if $S^3 = T$.*

19 Suppose $T \in \mathcal{L}(V)$ is self-adjoint, $\lambda \in \mathbf{F}$, and $\epsilon > 0$. Suppose there exists $v \in V$ such that $\|v\| = 1$ and

$$\|Tv - \lambda v\| < \epsilon.$$

Prove that T has an eigenvalue λ' such that $|\lambda - \lambda'| < \epsilon$.

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.

20 Give an alternative proof of the complex spectral theorem that avoids Schur's theorem and instead follows the pattern of the proof of the real spectral theorem.

21 Suppose U is a finite-dimensional vector space and $T \in \mathcal{L}(U)$.

(a) Suppose $\mathbf{F} = \mathbf{R}$. Prove that T is diagonalizable if and only if there is an inner product on U that makes T into a self-adjoint operator.

(b) Suppose $\mathbf{F} = \mathbf{C}$. Prove that T is diagonalizable if and only if there is an inner product on U that makes T into a normal operator.

This exercise adds another equivalence to the list of conditions equivalent to diagonalizability in 5.55.

22 Suppose that $T \in \mathcal{L}(V)$ 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 the pseudoinverse $T^\dagger$ is the operator on V such that

$$T^\dagger e_k = \begin{cases} \frac{1}{\lambda_k} e_k & \text{if } \lambda_k \neq 0, \\ 0 & \text{if } \lambda_k = 0. \end{cases}$$

7C *Positive Operators*

7.34 definition: *positive operator*

An operator $T \in \mathcal{L}(V)$ is called *positive* if T is self-adjoint and

$$\langle Tv, v\rangle \geq 0$$

for all $v \in V$.

If V is a complex vector space, then the requirement that T is self-adjoint can be dropped from the definition above (by 7.14).

7.35 example: *positive operators*

(a) If U is a subspace of V, then the orthogonal projection P_U is a positive operator, as you should verify.

(b) If $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbf{R}$ are such that $b^2 < 4c$, then $T^2 + bT + cI$ is a positive operator, as shown by the proof of 7.26.

7.36 definition: *square root*

An operator R is called a *square root* of an operator T if $R^2 = T$.

7.37 example: *square root of an operator*

If $T \in \mathcal{L}(\mathbf{F}^3)$ is defined by $T(z_1, z_2, z_3) = (z_3, 0, 0)$, then the operator $R \in \mathcal{L}(\mathbf{F}^3)$ defined by $R(z_1, z_2, z_3) = (z_2, z_3, 0)$ is a square root of T 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 $\mathbf{C}$. Specifically, a number $z \in \mathbf{C}$ 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 $w \in \mathbf{C}$ such that $z = \overline{w}w$, corresponding to condition (f). See Exercise 17 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.*

7.38 *characterization of positive operators*

Let $T \in \mathcal{L}(V)$. 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 $R \in \mathcal{L}(V)$

Proof We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (a).

First suppose (a) holds, so that T is positive. Obviously T is self-adjoint (by the definition of a positive operator). To prove the other condition in (b), suppose λ is an eigenvalue of T. Let v be an eigenvector of T corresponding to λ. Then

$$0 \leq \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle.$$

Thus λ is a nonnegative number. Hence (b) holds, showing that (a) implies (b).

Now suppose (b) holds, so that T is self-adjoint and all the eigenvalues of T are nonnegative. By the spectral theorem (7.29 and 7.31), there is an orthonormal basis $e_1, \ldots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of T corresponding to $e_1, \ldots, e_n$; thus each λ_k is a nonnegative number. The matrix of T with respect to $e_1, \ldots, e_n$ is the diagonal matrix with $\lambda_1, \ldots, \lambda_n$ on the diagonal, which shows that (b) implies (c).

Now suppose (c) holds. Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V such that the matrix of T with respect to this basis is a diagonal matrix with nonnegative numbers $\lambda_1, \ldots, \lambda_n$ on the diagonal. The linear map lemma (3.4) implies that there exists $R \in \mathcal{L}(V)$ such that

$$Re_k = \sqrt{\lambda_k}e_k$$

for each $k = 1, \ldots, n$. 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).

Clearly (d) implies (e) (because, by definition, every positive operator is self-adjoint).

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 $R \in \mathcal{L}(V)$ be such that $T = R^*R$. Then $T^* = (R^*R)^* = R^*(R^*)^* = R^*R = T$. Hence T is self-adjoint. To complete the proof that (a) holds, note that

$$\langle Tv, v\rangle = \langle R^*Rv, v\rangle = \langle Rv, Rv\rangle \geq 0$$

for every $v \in V$. Thus T is positive, showing that (f) implies (a). ■

Each nonnegative number has a unique nonnegative square root. The next result shows that positive operators enjoy a similar property.

7.39 *each positive operator has only one positive square root*

Every positive operator on V has a unique positive square root.

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 $\dim V > 1$.

Proof Suppose $T \in \mathcal{L}(V)$ is positive. Suppose $v \in V$ is an eigenvector of T. Hence there exists $\lambda \geq 0$ 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 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 $a_1, \dots, a_n \in \mathbf{F}$. Thus

$$Rv = a_1 \sqrt{\lambda_1} e_1 + \cdots + a_n \sqrt{\lambda_n} e_n$$

and 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

$$v = \sum_{\{k : \lambda_k = \lambda\}} a_k e_k,$$

and 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.

7.40 notation: $\sqrt{T}$

For T a positive operator, $\sqrt{T}$ denotes the unique positive square root of T.

7.41 example: *square root of positive operators*

Define operators S, T on $\mathbf{R}^2$ (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 $\mathbf{R}^2$ we have

7.42
$$\mathcal{M}(S) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathcal{M}(T) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Each of these matrices equals its transpose; thus S and T are self-adjoint.

If $(x, y) \in \mathbf{R}^2$, then

$$\langle S(x, y), (x, y) \rangle = x^2 + 2y^2 \geq 0$$

and

$$\langle T(x, y), (x, y) \rangle = x^2 + 2xy + y^2 = (x + y)^2 \geq 0.$$

Thus S and T are positive operators.

The standard basis of $\mathbf{R}^2$ is an orthonormal basis consisting of eigenvectors of S. Note that

$$\left(\tfrac{1}{\sqrt{2}}, \tfrac{1}{\sqrt{2}}\right), \left(\tfrac{1}{\sqrt{2}}, -\tfrac{1}{\sqrt{2}}\right)$$

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 with respect to the standard basis

$$\mathcal{M}\big(\sqrt{S}\,\big) = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{pmatrix} \quad \text{and} \quad \mathcal{M}\big(\sqrt{T}\,\big) = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

by showing that squares of the matrices above are the matrices in 7.42 and showing 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.

7.43 *T positive and $\langle Tv, v \rangle = 0 \implies Tv = 0$*

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 $T \in \mathcal{L}(\mathbf{F}^4)$ 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.

2 Prove that the sum of two positive operators on V is positive.

3 Suppose $T \in \mathcal{L}(V)$. Prove that T is a positive operator if and only if the pseudoinverse $T^\dagger$ is a positive operator.

4 Suppose $T \in \mathcal{L}(V)$ is a positive operator and $S \in \mathcal{L}(W, V)$. Prove that S^*TS is a positive operator on W.

5 Give an example of a self-adjoint operator T on some inner product space such that every entry of $\mathcal{M}(T)$ (with respect to some orthonormal basis) is positive, but T is not a positive operator.

6 Suppose n is a positive integer and $T \in \mathcal{L}(\mathbf{F}^n)$ is the operator whose matrix (with respect to the standard basis) consists of all 1's. Show that T is a positive operator.

7 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 $v = w$.

8 Suppose T is a positive operator on V and U is a subspace of V invariant under T. Prove that $T|_U \in \mathcal{L}(U)$ is a positive operator on U.

9 Suppose $T \in \mathcal{L}(V)$ is self-adjoint and $\alpha \in \mathbf{R}$.

(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.

10 Suppose $T \in \mathcal{L}(V)$ is positive. Prove that T^k is positive for every positive integer k.

11 Suppose T is a positive operator on V and $v_1, \ldots, v_m \in V$. Prove that

$$\sum_{j=1}^{m} \sum_{k=1}^{m} \langle Tv_k, v_j \rangle \geq 0.$$

12 Suppose $T \in \mathcal{L}(V)$ 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.

13 Suppose T is a positive operator on V. Prove that

$$\operatorname{null}\sqrt{T} = \operatorname{null} T \quad \text{and} \quad \operatorname{range}\sqrt{T} = \operatorname{range} T.$$

14 Suppose that $T \in \mathcal{L}(V)$ is a positive operator. Prove that there exists a polynomial $p \in \mathcal{P}(\mathbf{F})$ such that $\sqrt{T} = p(T)$.

15 Suppose S and T are positive operators on V. Prove that ST is a positive operator if and only if S and T commute.

16 Suppose T is a positive operator on V. Prove that

$$T \text{ is invertible} \iff \langle Tv, v\rangle > 0 \text{ for every } v \in V \text{ with } v \neq 0.$$

17 Suppose $T \in \mathcal{L}(V)$ and $e_1, \ldots, e_n$ is an orthonormal basis of V. Prove that T is a positive operator if and only if there exist $v_1, \ldots, v_n \in V$ such that

$$\langle Te_k, e_j\rangle = \langle v_k, v_j\rangle$$

for all $j, k = 1, \ldots, n$.

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 $e_1, \ldots, e_n$.

18 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 $\frac{1}{j+k-1}$. Suppose $T \in \mathcal{L}(V)$ 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

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{pmatrix}.$$

19 For $T \in \mathcal{L}(V)$ and $u, v \in V$, define $\langle u, v\rangle_T$ by $\langle u, v\rangle_T = \langle Tu, v\rangle$.

(a) Suppose $T \in \mathcal{L}(V)$. 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 each inner product on V is of the form $\langle \cdot, \cdot\rangle_T$ for some positive invertible operator $T \in \mathcal{L}(V)$.

20 Suppose S and T are positive operators on V. Prove that

$$\operatorname{null}(S + T) = \operatorname{null} S \cap \operatorname{null} T.$$

21 Prove or disprove: the identity operator on $\mathbf{F}^2$ has infinitely many self-adjoint square roots.

22 Let T be the second derivative operator in Exercise 28(b) in Section 7A. Show that $-T$ is a positive operator.

7D Isometries, Unitary Operators, and QR Factorization

Isometries

Linear maps that preserve norms are sufficiently important to deserve a name.

7.44 definition: *isometry*

A linear map $S \in \mathcal{L}(V, W)$ 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 $S \in \mathcal{L}(V, W)$ 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.*

7.45 example: *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 $S \in \mathcal{L}(V, W)$ 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

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \tag{7.46}$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2, \tag{7.47}$$

where we have used 6.30(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 $\|v\| = \|Sv\|$. In other words, 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 it maps some orthonormal basis to an orthonormal list. In other words, the isometries given by Example 7.45 include all the 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 choice of orthonormal bases) form an orthonormal list. Here we are identifying the columns of an m-by-n matrix with elements of $\mathbf{F}^m$ and then using the Euclidean inner product on $\mathbf{F}^m$.

7.49 *characterization of isometries*

Suppose $S \in \mathcal{L}(V, W)$. Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $f_1, \ldots, f_m$ is an orthonormal basis of W. Then 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 $\mathcal{M}(S, (e_1, \ldots, e_n), (f_1, \ldots, f_m))$ form an orthonormal list in $\mathbf{F}^m$ with respect to the Euclidean inner product

Proof First suppose (a) holds, so S is an isometry. If $v \in V$ then

$$\langle (I - S^*S)v, v \rangle = \langle v, v \rangle - \langle S^*Sv, v \rangle = \|v\|^2 - \langle Sv, Sv \rangle = \|v\|^2 - \|Sv\|^2 = 0.$$

Hence the self-adjoint operator $I - S^*S$ equals 0 (by 7.16). In other words, $S^*S = I$, 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 $j, k \in \{1, \ldots, n\}$, then

$$\langle Se_j, Se_k \rangle = \langle e_j, e_k \rangle.$$

Hence $Se_1, \ldots, Se_n$ is an orthonormal list in W, proving that (c) implies (d).

Now suppose that (d) holds, so $Se_1, \ldots, Se_n$ is an orthonormal list in W. Let $A = \mathcal{M}(S, (e_1, \ldots, e_n), (f_1, \ldots, f_m))$. For each $k, r \in \{1, \ldots, n\}$, we have

7.50 $$\sum_{j=1}^{m} A_{j,k} \overline{A_{j,r}} = \left\langle \sum_{j=1}^{m} A_{j,k} f_j, \sum_{j=1}^{m} A_{j,r} f_j \right\rangle = \langle Se_k, Se_r \rangle.$$

Thus the columns of A form an orthonormal list in $\mathbf{F}^m$, 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 $\mathbf{F}^n$. Then 7.50 shows that $Se_1, \ldots, Se_n$ is an orthonormal list in W. Thus Example 7.45, with $Se_1, \ldots, Se_n$ playing the role of $g_1, \ldots, g_n$, shows that S is an isometry, proving that (e) implies (a). ■

See Exercises 1 and 6 for additional conditions that are equivalent to being an isometry.

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.

7.51 definition: ***unitary operator***

An operator $S \in \mathcal{L}(V)$ is called *unitary* if S is an invertible isometry.

Although the words "unitary" and "isometry" mean the same thing when considering 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.

As previously noted, every isometry is injective. Every injective operator on a finite-dimensional vector space is invertible (see 3.54). 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.

7.52 example: *rotation of* $\mathbf{R}^2$

Suppose $\theta \in \mathbf{R}$ and S is the operator on $\mathbf{F}^2$ whose matrix with respect to the standard basis of $\mathbf{F}^2$ is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

The two columns of this matrix form an orthonormal list in $\mathbf{F}^2$; hence S is an isometry [by the equivalence of (a) and (e) in 7.49]. Thus S is a unitary operator.

If $\mathbf{F} = \mathbf{R}$, then S is the operator of counterclockwise rotation by θ radians around the origin of $\mathbf{R}^2$. This observation gives us another way to think about why S is an isometry, because each rotation around the origin of $\mathbf{R}^2$ 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 $S \in \mathcal{L}(V, W)$ is an isometry if and only if $S^*S = I$, while 7.53 shows that an operator $S \in \mathcal{L}(V)$ 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 $\mathcal{M}(T)$, while 7.53(e) mentions the rows of $\mathcal{M}(T)$. Furthermore, $\mathcal{M}(T)$ in 7.49(e) is with respect to an orthonormal basis of V and an orthonormal basis of W, while $\mathcal{M}(T)$ in 7.53(e) is with respect to a single basis of V doing double duty.

7.53 *characterization of unitary operators*

Suppose $S \in \mathcal{L}(V)$. 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 $\mathcal{M}(S, (e_1, \dots, e_n))$ form an orthonormal basis of $\mathbf{F}^n$ 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).

Clearly (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 last equation holds because we have already shown 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 $\mathcal{M}(S^*, (e_1, \dots, e_n))$ form an orthonormal basis of $\mathbf{F}^n$ [by the equivalence of (a) and (e) of 7.49]. The rows of $\mathcal{M}(S, (e_1, \dots, e_n))$ are the complex conjugates of the columns of $\mathcal{M}(S^*, (e_1, \dots, e_n))$. Thus the rows of $\mathcal{M}(S, (e_1, \dots, e_n))$ form an orthonormal basis of $\mathbf{F}^n$, proving that (d) implies (e).

Now suppose (e) holds. Thus the columns of $\mathcal{M}(S^*, (e_1, \dots, e_n))$ form an orthonormal basis of $\mathbf{F}^n$. 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 (a) ⇒ (b) ⇒ (c) ⇒ (d) ⇒ (e) ⇒ (f) ⇒ (a), completing the proof. ■

Recall our analogy between $\mathbf{C}$ and $\mathcal{L}(V)$. Under this analogy, a complex number z corresponds to an operator $S \in \mathcal{L}(V)$, and $\bar{z}$ corresponds to S^*. The real numbers ($z = \bar{z}$) correspond to the self-adjoint operators ($S = S^*$), and the nonnegative numbers correspond to the (badly named) positive operators.

Another distinguished subset of $\mathbf{C}$ is the unit circle, which consists of the complex numbers z such that $|z| = 1$. The condition $|z| = 1$ is equivalent to the condition $\bar{z}z = 1$. Under our analogy, this would correspond to the condition $S^*S = I$, which is equivalent to S being a unitary operator. In other words, the analogy shows that the unit circle in $\mathbf{C}$ corresponds to the set of unitary operators. In the next two results, this analogy appears in the eigenvalues of unitary operators. Also see Exercise 11 for another example of this analogy.

7.54 *eigenvalues of unitary operators have absolute value* 1

Suppose λ is an eigenvalue of a unitary operator. Then $|\lambda| = 1$.

Proof Suppose $S \in \mathcal{L}(V)$ 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.

7.55 *description of unitary operators on complex inner product spaces*

Suppose $\mathbf{F} = \mathbf{C}$ and $S \in \mathcal{L}(V)$. 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. Let $g_k = \lambda_k e_k$ for each $k = 1, \dots, n$. Then $g_1, \dots, g_n$ is also an orthonormal basis of V because

$$\langle g_j, g_k \rangle = \langle \lambda_j e_j, \lambda_k e_k \rangle = \lambda_j \overline{\lambda_k} \langle e_j, e_k \rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k \end{cases}$$

for all $j, k = 1, \dots, n$. Furthermore, $Se_k = g_k$ for each k. Thus Example 7.45 shows that S is an isometry, 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 $\mathbf{F}^n$ (with the Euclidean inner product) and that their matrices are computed with respect to the standard basis of $\mathbf{F}^n$.

We begin by making the following definition, transferring the notion of a unitary operator to a unitary matrix.

7.56 definition: ***unitary matrix***

An n-by-n matrix is called *unitary* if its columns form an orthonormal list in $\mathbf{F}^n$.

In the definition above, we could have replaced “orthonormal list in $\mathbf{F}^n$” with “orthonormal basis of $\mathbf{F}^n$” because every orthonormal list of length n in an n-dimensional inner product space is an orthonormal basis. If $S \in \mathcal{L}(V)$ and $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ are orthonormal bases of V, then S is a unitary operator if and only if $\mathcal{M}(S, (e_1, \ldots, e_n), (f_1, \ldots, f_n))$ 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 part (c), Qv denotes the matrix product of Q and v, identifying elements of $\mathbf{F}^n$ with n-by-1 matrices (sometimes called column vectors). The norm in part (c) below is the usual Euclidean norm on $\mathbf{F}^n$ that comes from the Euclidean inner product. In part (d), Q^* denotes the conjugate transpose of the matrix Q, which corresponds to the adjoint of the associated operator.

7.57 ***characterizations of unitary matrices***

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 $\mathbf{F}^n$

(c) $\|Qv\| = \|v\|$ for every $v \in \mathbf{F}^n$

(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.

7.58 ***QR factorization***

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 all positive entries on its diagonal, and

$$A = QR.$$

Proof Let $v_1, \ldots, v_n$ denote the columns of A, thought of as elements of $\mathbf{F}^n$. Apply the Gram–Schmidt procedure (6.32) to the list $v_1, \ldots, v_n$, getting an orthonormal basis $e_1, \ldots, e_n$ of $\mathbf{F}^n$ such that

7.59 $$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k)$$

for each $k = 1, \ldots, n$. 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 $\operatorname{span}(e_1, \ldots, e_k)$ 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 $e_1, \ldots, e_n$. If $k \in \{1, \ldots, n\}$, then the k^{th} column of QR equals a linear combination of the columns of Q, with the coefficients for the linear combination coming from the k^{th} column of R (see 3.45 and 3.47). In other words, the k^{th} 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 6.30(a)], the k^{th} column of A. Thus $A = QR$, as desired.

The equations defining the Gram–Schmidt procedure (see 6.32) show that each v_k equals a positive multiple of e_k plus a linear combination of $e_1, \ldots, e_{k-1}$. 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 all positive entries on its diagonal. Let $q_1, \ldots, q_n$ denote the columns of $\widehat{Q}$. Thinking of matrix multiplication as above, we see that each v_k is a linear combination of $q_1, \ldots, q_k$, with the coefficients coming from the k^{th} column of $\widehat{R}$. This implies that $\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(q_1, \ldots, q_k)$ 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 $k = 1, \ldots, n$. Hence $\widehat{Q} = Q$, which then implies that $\widehat{R} = R$, completing the proof of uniqueness. ■

For a nice application of the QR factorization, see the proof of Hadamard's inequality (10.49).

The proof of the QR factorization shows that the columns of the unitary matrix Q can be computed by applying the Gram–Schmidt procedure to the columns of the input matrix A. The next example illustrates the computation of the QR factorization based the proof of 7.58.

7.60 example: *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

$$e_1 = (1,0,0), \quad e_2 = \left(0, \tfrac{1}{\sqrt{10}}, -\tfrac{3}{\sqrt{10}}\right), \quad e_3 = \left(0, \tfrac{3}{\sqrt{10}}, \tfrac{1}{\sqrt{10}}\right).$$

Still following the proof of 7.58, let Q be the unitary matrix whose columns are e_1, e_2, e_3:

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ 0 & \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}.$$

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

$$R = \begin{pmatrix} 1 & 2 & 1 \\ 0 & \sqrt{10} & \frac{\sqrt{10}}{5} \\ 0 & 0 & \frac{7\sqrt{10}}{5} \end{pmatrix}.$$

Note that R is indeed an upper-triangular matrix with positive entries on the diagonal, as required by the QR factorization.

Now matrix multiplication can verify that $A = QR$ is the desired factorization of A:

$$QR = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \\ 0 & \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & \sqrt{10} & \frac{\sqrt{10}}{5} \\ 0 & 0 & \frac{7\sqrt{10}}{5} \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -4 \\ 0 & 3 & 2 \end{pmatrix} = A.$$

Thus $A = QR$, as expected.

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 $b \in \mathbf{F}^n$ and we want to solve the equation $Ax = b$ for $x = (x_1, \dots, x_n) \in \mathbf{F}^n$ (as usual, we are identifying elements of $\mathbf{F}^n$ with n-by-1 column vectors).

Suppose $A = QR$, where Q is unitary and R is upper triangular with all positive entries 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 easy to compute because it is just the conjugate transpose of the matrix Q. Because R is an upper-triangular matrix with positive entries on its diagonal, the system of linear equations represented by the equation above can easily be solved by first solving for x_n, then for x_{n-1}, and so on.

Exercises 7D

1 Suppose $\dim V \geq 2$ and $S \in \mathcal{L}(V, W)$. 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 2 in V.

2 Suppose $S \in \mathcal{L}(V, W)$ preserves orthogonality, meaning that $u, v \in V$ and $\langle u, v \rangle = 0$ implies $\langle Su, Sv \rangle = 0$. Suppose also that there exists $e \in V$ such that $e \neq 0$ and $\|Se\| = \|e\|$. Prove that S is an isometry.

3 Suppose T_1, T_2 are normal operators on $\mathbf{F}^3$ and both operators have $2, 5, 7$ as eigenvalues. Prove that there exists a unitary operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T_1 = S^*T_2S$.

4 Give an example of two self-adjoint operators $T_1, T_2 \in \mathcal{L}(\mathbf{F}^4)$ such that the eigenvalues of both operators are $2, 5, 7$ but there does not exist a unitary operator $S \in \mathcal{L}(\mathbf{F}^4)$ such that $T_1 = S^*T_2S$. Be sure to explain why there is no unitary operator with the required property.

5 Prove or give counterexample: If $S \in \mathcal{L}(V)$ 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.

6 Suppose $S \in \mathcal{L}(V)$. Prove that S is a unitary operator if and only if

$$\{Sv : v \in V \text{ and } \|v\| \leq 1\} = \{v \in V : \|v\| \leq 1\}.$$

7 Prove or give counterexample: If $S \in \mathcal{L}(V)$ is invertible and $\|S^{-1}v\| = \|Sv\|$ for every $v \in V$, then S is unitary.

8 Explain why the columns of a square matrix of complex numbers form an orthonormal list in $\mathbf{C}^n$ if and only if the rows of the matrix form an orthonormal list in $\mathbf{C}^n$.

9 Suppose $v \in V$ with $\|v\| = 1$ and $b \in \mathbf{F}$. Also suppose $\dim V \geq 2$. Prove there exists a unitary operator $S \in \mathcal{L}(V)$ such that $\langle Sv, v \rangle = b$ if and only if $|b| \leq 1$.

10 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Suppose every eigenvalue of T has absolute value 1 and $\|Tv\| \leq \|v\|$ for every $v \in V$. Prove that T is a unitary operator.

11 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 $\mathbf{F} = \mathbf{C}$, 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 $\mathbf{C}$ (except for the point 1) to $\mathbf{R}$. Thus part (b) illustrates the analogy between the unitary operators and the unit circle in $\mathbf{C}$, along with the analogy between the self-adjoint operators and $\mathbf{R}$.

12 Explain why the characterization of unitary matrices given by 7.57 holds.

7E Singular Value Decomposition

Singular Values

We will need to use the following result in this section.

7.61 *properties of T^*T*

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

(a) T^*T is a positive operator on V;

(b) $\operatorname{null} T^*T = \operatorname{null} T$;

(c) $\operatorname{range} T^*T = \operatorname{range} T^*$;

(d) $\dim \operatorname{range} T = \dim \operatorname{range} T^* = \dim \operatorname{range} T^*T$.

Proof

(a) We have

$$(T^*T)^* = T^*(T^*)^* = T^*T.$$

Thus T^*T is self-adjoint.

If $v \in V$, then

$$\langle (T^*T)v, v \rangle = \langle T^*(Tv), v \rangle = \langle Tv, Tv \rangle = \|Tv\|^2 \geq 0.$$

Thus T^*T is a positive operator.

(b) First suppose $v \in \operatorname{null} T^*T$. Then

$$\|Tv\|^2 = \langle Tv, Tv \rangle = \langle T^*Tv, v \rangle = \langle 0, v \rangle = 0.$$

Thus $Tv = 0$, proving that $\operatorname{null} T^*T \subset \operatorname{null} T$.

The inclusion in the other direction is clear, because if $v \in V$ and $Tv = 0$, then $T^*Tv = 0$.

Thus $\operatorname{null} T^*T = \operatorname{null} T$, completing the proof of (b).

(c) We already know from part (a) that T^*T is self-adjoint. Thus

$$\operatorname{range} T^*T = (\operatorname{null} T^*T)^\perp = (\operatorname{null} T)^\perp = \operatorname{range} T^*,$$

where the first and last equalities come from 7.6 and the second equality comes from part (b).

(d) To verify the first equation in part (d), note that

$$\dim \operatorname{range} T = \dim (\operatorname{null} T^*)^\perp = \dim W - \dim \operatorname{null} T^* = \dim \operatorname{range} T^*,$$

where the first equality comes from 7.6(d), the second equality comes from 6.53, and the last equality comes from the fundamental theorem of linear maps (3.21).

The equality $\dim \operatorname{range} T^* = \dim \operatorname{range} T^*T$ follows from part (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.

7.62 definition: *singular values*

Suppose $T \in \mathcal{L}(V, W)$. 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.

7.63 example: *singular values of an operator on* $\mathbf{F}^4$

Define $T \in \mathcal{L}(\mathbf{F}^4)$ by $T(z_1, z_2, z_3, z_4) = (0, 3z_1, 2z_2, -3z_4)$. 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 $\mathbf{F}^4$ diagonalizes T^*T, and we see that the eigenvalues of T^*T are 9, 4, and 0. Also, the dimensions of the 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.

7.64 example: *singular values of a linear map from* $\mathbf{F}^4$ *to* $\mathbf{F}^3$

Suppose $T \in \mathcal{L}(\mathbf{F}^4, \mathbf{F}^3)$ 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.

7.65 *role of positive singular values*

Suppose that $T \in \mathcal{L}(V, W)$. Then

(a) T is injective $\iff$ 0 is not a singular value of T;

(b) the number of positive singular values of T equals $\dim \operatorname{range} T$;

(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 $\operatorname{null} T = \{0\}$, which happens if and only if $\operatorname{null} T^*T = \{0\}$ [by 7.61(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 $\dim \operatorname{range} T^*T$ equals the number of positive eigenvalues of T^*T (counting repetitions). Thus 7.61(c) implies that $\dim \operatorname{range} T$ equals the number of positive singular values of T, proving (b).

Clearly (b) implies (c). ■

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 $\mathbf{F} = \mathbf{R}$) or complex numbers (if $\mathbf{F} = \mathbf{C}$)	are nonnegative numbers
can be the empty list if $\mathbf{F} = \mathbf{R}$	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 $\mathbf{F} = \mathbf{C}$	always listed in decreasing order

The next result nicely characterizes isometries in terms of singular values.

7.66 *isometries characterized by having all singular values equal* 1

Suppose that $S \in \mathcal{L}(V, W)$. Then

$$S \text{ is an isometry} \iff \text{all the singular values of } S \text{ equal } 1.$$

Proof We have

$$\begin{aligned} S \text{ is an isometry} &\iff S^*S = I \\ &\iff \text{all the eigenvalues of } S^*S \text{ equal } 1 \\ &\iff \text{all the singular values of } S \text{ equal } 1, \end{aligned}$$

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. ■

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. We will see several important applications of the singular value decomposition (often called the SVD). In addition, the singular value decomposition is useful in computational linear algebra because good techniques exist for approximating eigenvalues and eigenvectors of positive operators such as the operator T^*T, whose eigenvalues and eigenvectors lead to the singular value decomposition.

7.67 *singular value decomposition*

Suppose $T \in \mathcal{L}(V, W)$ 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

7.68 $$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

for every $v \in V$.

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 7.61(a)], the spectral theorem implies that there exists an orthonormal basis $e_1, \dots, e_n$ of V with

7.69 $$T^*Te_k = s_k^2 e_k$$

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

For each $k = 1, \dots, m$, let

7.70 $$f_k = \frac{Te_k}{s_k}.$$

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 7.69), which implies that $Te_k = 0$ [by 7.61(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 7.70. The equation above is our desired result. ■

Suppose $T \in \mathcal{L}(V, W)$, the positive singular values of T are $s_1, \ldots, s_m$, and $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ are as in the singular value decomposition 7.67. The orthonormal list $e_1, \ldots, e_m$ can be extended to an orthonormal basis $e_1, \ldots, e_{\dim V}$ of V and the orthonormal list $f_1, \ldots, f_m$ can be extended to an orthonormal basis $f_1, \ldots, f_{\dim W}$ of W. The formula 7.68 shows that

$$Te_k = \begin{cases} s_k f_k & \text{if } 1 \le k \le m, \\ 0 & \text{if } m < k \le \dim V. \end{cases}$$

Thus the matrix of T with respect to the orthonormal bases $(e_1, \ldots, e_{\dim V})$ and $(f_1, \ldots, f_{\dim W})$ has the simple form

$$\mathcal{M}\big(T, (e_1, \ldots, e_{\dim V}), (f_1, \ldots, f_{\dim W})\big)_{j,k} = \begin{cases} s_k & \text{if } 1 \le j = k \le m, \\ 0 & \text{otherwise.} \end{cases}$$

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 shown the wonderful result that every linear map from V to W has a diagonal matrix with respect to appropriate orthonormal bases.

7.71 definition: ***diagonal matrix***

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 $k = 1, \ldots, \min\{M, N\}$.

The table below compares the spectral theorem (7.29 and 7.31) with the singular value decomposition (7.67).

spectral theorem	**singular value decomposition**
describes only self-adjoint operators (when $\mathbf{F} = \mathbf{R}$) or normal operators (when $\mathbf{F} = \mathbf{C}$)	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 upon whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$	same proof works regardless of whether $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$

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 $T \in \mathcal{L}(V, W)$, we can obtain the adjoint of T simply by interchanging the role of the e's and the f's (see 7.73 and 7.74). Similarly, we can obtain the pseudoinverse $T^\dagger$ (see 6.70) of T by interchanging the role of the e's and the f's and replacing each positive singular value s_k of T with $1/s_k$ (see 7.73 and 7.75).

Recall that the pseudoinverse $T^\dagger$ in 7.75 below equals the inverse T^{-1} if T is invertible [see 6.71(a)].

7.72 *singular value decomposition of adjoint and pseudoinverse*

Suppose $T \in \mathcal{L}(V, W)$ and the positive singular values of T are $s_1, \ldots, s_m$. Suppose $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ are orthonormal lists in V and W such that

7.73
$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for every $v \in V$. Then

7.74
$$T^*w = s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m$$

and

7.75
$$T^\dagger w = \frac{\langle w, f_1\rangle}{s_1}e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m}e_m$$

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 7.74.

To prove 7.75, 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

$$\begin{aligned}Tv &= \frac{\langle w, f_1\rangle}{s_1}Te_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m}Te_m \\ &= \langle w, f_1\rangle f_1 + \cdots + \langle w, f_m\rangle f_m \\ &= w,\end{aligned}$$

where the second line holds because 7.73 implies that $Te_k = s_k f_k$ if $k = 1, \ldots, m$, and the third line above holds because 7.73 implies that $f_1, \ldots, f_m$ spans range T and thus is an orthonormal basis of range T. The equation above and the observation that $v \in (\operatorname{null} T)^\perp$ [see Exercise 8(b)] show $v = T^\dagger w$, proving 7.75. ∎

7.76 example: *finding a singular value decomposition*

Define $T \in \mathcal{L}(\mathbf{F}^4, \mathbf{F}^3)$ 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 7.64, 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 $\mathbf{F}^4$ and an orthonormal list f_1, f_2 in $\mathbf{F}^3$ such that

$$Tv = 5\langle v, e_1\rangle f_1 + \sqrt{2}\langle v, e_2\rangle f_2$$

for all $v \in \mathbf{F}^4$.

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 $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right)$. Thus, following the proof of 7.67, we take

$$e_1 = (0, 0, 0, 1) \quad \text{and} \quad e_2 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0\right)$$

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 $\mathbf{F}^4$ and f_1, f_2 is an orthonormal list in $\mathbf{F}^3$ and

$$Tv = 5\langle v, e_1\rangle f_1 + \sqrt{2}\langle v, e_2\rangle f_2$$

for all $v \in \mathbf{F}^4$. 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 "orthogonal columns" should be interpreted to mean that the columns are orthogonal with respect to the standard Euclidean inner product.

7.77 *matrix version of SVD*

Suppose A is an M-by-n matrix with rank $m \geq 1$. Then there exist an M-by-m matrix B with orthonormal columns, an m-by-m diagonal matrix D with positive entries on the diagonal, and an n-by-m matrix C with orthonormal columns such that

$$A = BDC^*.$$

Proof Let $T\colon \mathbf{F}^n \to \mathbf{F}^M$ be the linear map whose matrix with the standard bases equals A. Then $\dim \operatorname{range} T = m$ (by 3.117). Let

7.78 $$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

be a singular value decomposition of T. Let

$B =$ the M-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 $\mathbf{F}^m$. 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 C) on the right to get

$$ACC^* = BDC^*.$$

Note that the rows of C^* are the complex conjugates of $e_1, \ldots, e_m$. Thus if $k \in \{1, \ldots, m\}$, 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 $v \in \operatorname{span}(e_1, \ldots, e_m)$.

If $v \in (\operatorname{span}(e_1, \ldots, e_m))^\perp$, then $Av = 0$ (as follows from 7.78) and $C^*v = 0$ (as follows from the definition of matrix multiplication). Hence $ACC^*v = Av$ for all $v \in (\operatorname{span}(e_1, \ldots, e_m))^\perp$.

Because ACC^* and A agree on $\operatorname{span}(e_1, \ldots, e_m)$ and on $(\operatorname{span}(e_1, \ldots, e_m))^\perp$, 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 Mn entries. In comparison, the matrices B, D, and C above have a total of

$$m(M + m + n)$$

entries. Thus if M and n are large numbers and the rank m is considerably less than M and n, then the number of entries that must be stored on a computer to represent A is considerably less than Mn.

Exercises 7E

1 Suppose $T \in \mathcal{L}(V, W)$. Show that $T = 0$ if and only if all the singular values of T equal 0.

2 Suppose $T \in \mathcal{L}(V, W)$ 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.*

3 Give an example of $T \in \mathcal{L}(\mathbf{C}^2)$ such that 0 is the only eigenvalue of T and the singular values of T are $5, 0$.

4 Suppose that $T \in \mathcal{L}(V, W)$, s_1 is the largest singular value of T, and s_n is the smallest singular value of T. Prove that

$$\{\|Tv\| : v \in V \text{ and } \|v\| = 1\} = [s_n, s_1].$$

5 Suppose $T \in \mathcal{L}(\mathbf{C}^2)$ is defined by $T(x, y) = (-4y, x)$. Find the singular values of T.

6 Find the singular values of the differentiation operator $D \in \mathcal{L}(\mathcal{P}_2(\mathbf{R}))$ defined by $Dp = p'$, where the inner product on $\mathcal{P}_2(\mathbf{R})$ is as in Example 6.34.

7 Suppose that $T \in \mathcal{L}(V)$ is self-adjoint or that $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ 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 are sorted into decreasing order.

8 Suppose $T \in \mathcal{L}(V, W)$. 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 range T.
(b) Prove that $e_1, \ldots, e_m$ is an orthonormal basis of $(\operatorname{null} T)^\perp$.
(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 if $k \in \{1, \ldots, m\}$, then f_k is an eigenvector of TT^* with corresponding eigenvalue s_k^2.

9 Suppose $T \in \mathcal{L}(V, W)$. Show that T and T^* have the same positive singular values.

10 Suppose $T \in \mathcal{L}(V, W)$ has singular values $s_1, \ldots, s_n$. Prove that if T is an invertible linear map, then T^{-1} has singular values

$$\frac{1}{s_n}, \ldots, \frac{1}{s_1}.$$

11 Suppose that $T \in \mathcal{L}(V, W)$ and $v_1, \ldots, v_n$ is an orthonormal basis of V. Prove that

$$\|Tv_1\|^2 + \cdots + \|Tv_n\|^2 = s_1^2 + \cdots + s_n^2,$$

where $s_1, \ldots, s_n$ are the singular values of T.

See the comment after Exercise 5 in Section 7A.

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 $T \in \mathcal{L}(V)$ is normal. Prove that the singular values of T^2 equal the squares of the singular values of T.

13 Suppose $T_1, T_2 \in \mathcal{L}(V)$. Prove that T_1 and T_2 have the same singular values if and only if there exist unitary operators $S_1, S_2 \in \mathcal{L}(V)$ such that $T_1 = S_1 T_2 S_2$.

14 Suppose $T \in \mathcal{L}(V, W)$. Let s_n denote the smallest singular value of T. Prove that $s_n\|v\| \leq \|Tv\|$ for every $v \in V$.

15 Suppose $T \in \mathcal{L}(V)$ and $s_1 \geq \cdots \geq s_n$ are the singular values of T. Prove that if λ is an eigenvalue of T, then $s_1 \geq |\lambda| \geq s_n$.

16 Suppose $T \in \mathcal{L}(V, W)$. Prove that $(T^*)^\dagger = (T^\dagger)^*$.

Compare the result in this exercise to the analogous result for invertible linear maps [*see* 7.5(*f*)]

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 $\|Tv\|$.

7.79 *upper bound for $\|Tv\|$*

Suppose $T \in \mathcal{L}(V, W)$. Let s_1 be the largest singular value of T. Then

$$\|Tv\| \leq s_1 \|v\|$$

for all $v \in V$.

For a lower bound on $\|Tv\|$, look at Exercise 14 in Section 7E.

Proof Let $s_1, \ldots, s_m$ denote the positive singular values of T, and let $e_1, \ldots, e_m$ be an orthonormal list in V and $f_1, \ldots, f_m$ be an orthonormal list in W that provide a singular value decomposition of T. In other words, if $v \in V$, then

7.80 $$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m.$$

Hence if $v \in V$ then

$$\begin{aligned} \|Tv\|^2 &= s_1^2 |\langle v, e_1 \rangle|^2 + \cdots + s_m^2 |\langle v, e_m \rangle|^2 \\ &\leq s_1^2 \big(|\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2 \big) \\ &\leq s_1^2 \|v\|^2, \end{aligned}$$

where the last inequality follows from Bessel's inequality (6.26). Taking square roots of both sides of the inequality above shows that $\|Tv\| \leq s_1 \|v\|$, as desired. ■

Suppose $T \in \mathcal{L}(V, W)$ and s_1 is the largest singular value of T. The result above shows that

7.81 $$\|Tv\| \leq s_1 \text{ for all } v \in V \text{ with } \|v\| \leq 1.$$

Taking $v = e_1$ in 7.80 shows that $Te_1 = s_1 f_1$. Because $\|f_1\| = 1$, this implies that $\|Te_1\| = s_1$. Thus because $\|e_1\| = 1$, 7.81 leads to the equation

7.82 $$\max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\} = s_1.$$

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.

7.83 definition: *norm of a linear map*, $\|\cdot\|$

Suppose $T \in \mathcal{L}(V, W)$. Then the *norm* of T, denoted $\|T\|$, is defined by

$$\|T\| = \max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\}.$$

In general, the maximum of an infinite set of nonnegative numbers need not exist. However, the discussion before 7.83 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 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 $T \in \mathcal{L}(V, W)$. The norm $\|T\|$ for $T \in \mathcal{L}(V, W)$ does not usually come from taking an inner product of T with itself (see Exercise 15). You should be able to tell from the context and from the symbols used which meaning of the norm is intended.

The properties below of the norm on $\mathcal{L}(V, W)$ look identical to properties of the norm on an inner product space (see 6.9 and 6.17). Part (d) below is called the *triangle inequality*. For the reverse triangle inequality, see Exercise 1.

7.84 *basic properties of norms of linear maps*

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

(a) $\|T\| \geq 0$;

(b) $\|T\| = 0 \iff T = 0$;

(c) $\|\lambda T\| = |\lambda|\, \|T\|$ for all $\lambda \in \mathbf{F}$;

(d) $\|S + T\| \leq \|S\| + \|T\|$ for all $S \in \mathcal{L}(V, W)$.

Proof

(a) The definition of $\|T\|$ clearly 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

$$Tu = \|u\|\, T\left(\frac{u}{\|u\|}\right) = 0,$$

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 $\lambda \in \mathbf{F}$. 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 $S \in \mathcal{L}(V, W)$, $v \in V$, $\|v\| \leq 1$ and $\|S + T\| = \|(S + T)v\|$. Then

$$\|S + T\| = \|(S + T)v\| = \|Sv + Tv\| \leq \|Sv\| + \|Tv\| \leq \|S\| + \|T\|,$$

competing the proof of (d). ■

For $S, T \in \mathcal{L}(V, W)$, 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 12 asserts that for every $T \in \mathcal{L}(V)$, there is an invertible operator as close to T as we wish.

7.85 *alternative formulas for* $\|T\|$

Suppose $V \neq 0$ and $T \in \mathcal{L}(V, W)$. 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 7.82.

(b) Let $v \in V$ be such that $0 < \|v\| \leq 1$. Let $u = v/\|v\|$. Then

$$\|u\| = \Big\|\frac{v}{\|v\|}\Big\| = 1 \quad \text{and} \quad \|Tu\| = \Big\|T\Big(\frac{v}{\|v\|}\Big)\Big\| = \frac{\|Tv\|}{\|v\|} \geq \|Tv\|.$$

Thus when finding the maximum of $\|Tv\|$ with $\|v\| \leq 1$, we can restrict attention to vectors in V with norm 1, proving (b).

(c) Suppose $v \in V$ and $v \neq 0$. Then the definition of $\|T\|$ implies that

$$\Big\|T\Big(\frac{v}{\|v\|}\Big)\Big\| \leq \|T\|,$$

which implies that

7.86 $$\|Tv\| \leq \|T\|\,\|v\|.$$

Now suppose $c \geq 0$ and $\|Tv\| \leq c\|v\|$ for all $v \in V$. This implies that

$$\|Tv\| \leq c \text{ for all } v \in V \text{ with } \|v\| \leq 1.$$

Taking the maximum of the left side of the inequality above over all $v \in V$ with $\|v\| \leq 1$ shows that $\|T\| \leq c$. Thus $\|T\|$ is the smallest number c such that $\|Tv\| \leq c\|v\|$ for all $v \in V$. ■

The equation 7.86 is frequently used when working with norms.

For computing an approximation of the norm of a linear map T given the matrix of T with respect to some orthonormal bases, 7.85(a) is likely to be most useful. The matrix of T^*T is easily computable from matrix multiplication. Then a computer can be asked to find an approximation for the largest eigenvalue of T^*T (excellent numeric algorithms exist for this purpose). Then taking the square root and using 7.85(a) gives an approximation for the norm of T (which usually cannot be computed exactly).

You should verify all the assertions in the example below.

7.87 example: *norms*

- If I denotes the usual identity operator on V, then $\|I\| = 1$.
- If $T \in \mathcal{L}(\mathbf{F}^n)$ and the matrix of T with respect to the usual basis of $\mathbf{F}^n$ consists of all 1's, then $\|T\| = n$.
- If $T \in \mathcal{L}(V)$ 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 $|\lambda_1|, \dots, |\lambda_n|$.
- If $T \in \mathcal{L}(\mathbf{R}^5)$ 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 $1/(j^2 + k)$, then standard mathematical software shows that the square root of the largest eigenvalue of T^*T is approximately 0.811. Thus $\|T\| \approx 0.811$. It is not possible to find an exact formula for the norm of this operator.

A linear map and its adjoint have the same norm, as shown by the next result.

7.88 *norm of the adjoint*

Suppose $T \in \mathcal{L}(V, W)$. Then $\|T^*\| = \|T\|$.

Proof Suppose $w \in V$. Then

$$\|T^*w\|^2 = \langle T^*w, T^*w \rangle = \langle TT^*w, w \rangle \le \|TT^*w\| \, \|w\| \le \|T\| \, \|T^*w\| \, \|w\|.$$

Dividing both sides of the inequality above by $\|T^*w\|$ gives the inequality

$$\|T^*w\| \le \|T\| \, \|w\|,$$

which along with 7.85(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 by 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 find the best approximation to 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 $\dim \operatorname{range} T_k = k$ and T_k minimizes the distance to T among all linear maps whose range has dimension at most k. This result leads to algorithms for compressing huge matrices while preserving their most important information.

7.89 *best approximation by linear map whose range has dimension* $\leq k$

Suppose $T \in \mathcal{L}(V, W)$ and $s_1 \geq \cdots \geq s_m$ are the positive singular values of T. Suppose $1 \leq k < m$. Then

$$\min\{\|T - S\| : S \in \mathcal{L}(V, W) \text{ and } \dim \operatorname{range} S \leq k\} = s_{k+1}.$$

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 $T_k \in \mathcal{L}(V, W)$ 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 $\dim \operatorname{range} T_k = k$ and $\|T - T_k\| = s_{k+1}$.

Proof If $v \in V$ then

$$\begin{aligned}\|(T - T_k)v\|^2 &= \|s_{k+1} \langle v, e_{k+1} \rangle f_{k+1} + \cdots + s_m \langle v, e_m \rangle f_m\|^2 \\ &= {s_{k+1}}^2 |\langle v, e_{k+1} \rangle|^2 + \cdots + {s_m}^2 |\langle v, e_m \rangle|^2 \\ &\leq {s_{k+1}}^2 \big(|\langle v, e_{k+1} \rangle|^2 + \cdots + |\langle v, e_m \rangle|^2\big) \\ &\leq {s_{k+1}}^2 \|v\|^2.\end{aligned}$$

Thus $\|T - T_k\| \leq s_{k+1}$. The equation $(T - T_k)e_{k+1} = s_{k+1} f_{k+1}$ now shows that $\|T - T_k\| = s_{k+1}$.

Suppose $S \in \mathcal{L}(V, W)$ and $\dim \operatorname{range} S \leq k$. Thus $Se_1, \ldots, Se_{k+1}$, which is a list of length $k + 1$, is linearly dependent. Hence there exist $a_1, \ldots, a_{k+1} \in \mathbf{F}$, not all 0, such that

$$a_1 Se_1 + \cdots + a_{k+1} Se_{k+1} = 0.$$

Now $a_1 e_1 + \cdots + a_{k+1} e_{k+1} \neq 0$ because $a_1, \ldots, a_{k+1}$ are not all 0. We have

$$\begin{aligned}\|(T - S)(a_1 e_1 + \cdots + a_{k+1} e_{k+1})\|^2 &= \|T(a_1 e_1 + \cdots + a_{k+1} e_{k+1})\|^2 \\ &= \|s_1 a_1 f_1 + \cdots + s_{k+1} a_{k+1} f_{k+1}\|^2 \\ &= s_1^2 |a_1|^2 + \cdots + {s_{k+1}}^2 |a_{k+1}|^2 \\ &\geq {s_{k+1}}^2 (|a_1|^2 + \cdots + |a_{k+1}|^2) \\ &= {s_{k+1}}^2 \|a_1 e_1 + \cdots + a_{k+1} e_{k+1}\|^2.\end{aligned}$$

Because $a_1 e_1 + \cdots + a_{k+1} e_{k+1} \neq 0$, the inequality above implies that

$$\|T - S\| \geq s_{k+1}.$$

Thus $S = T_k$ minimizes $\|T - S\|$ among $S \in \mathcal{L}(V, W)$ with $\dim \operatorname{range} S \leq k$. ■

For other examples of the use of the singular value decomposition in best approximation, see Exercise 16, which finds a subspace of given dimension on which the restriction of a linear map is as small as possible, and Exercise 17, 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 each complex number z except 0 can be written in the form

$$\begin{aligned} z &= \left(\frac{z}{|z|}\right)|z| \\ &= \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z}, \end{aligned}$$

where the first factor, namely, $z/|z|$, has absolute value 1.

Our analogy leads us to guess that each operator $T \in \mathcal{L}(V)$ 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 $T \in \mathcal{L}(V)$, then T^*T is a positive operator [as was shown in 7.61(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 each 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 $\mathbf{F} = \mathbf{C}$; the positive operators are described by the real and complex spectral theorems (7.29 and 7.31).

Specifically, consider the case $\mathbf{F} = \mathbf{C}$, and suppose

$$T = S\sqrt{T^*T}$$

is a polar decomposition of an operator $T \in \mathcal{L}(V)$, where S is 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. In other words, S may require one orthonormal basis and $\sqrt{T^*T}$ may require a different orthonormal basis.

However, (still assuming that $\mathbf{F} = \mathbf{C}$) 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 26. The converse is also true: If $T \in \mathcal{L}(V)$ and $T = S\sqrt{T^*T}$ for some unitary operator $S \in \mathcal{L}(V)$ 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.

7.90 *polar decomposition*

Suppose $T \in \mathcal{L}(V)$. Then there exists a unitary operator $S \in \mathcal{L}(V)$ 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

7.91 $$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

for every $v \in V$. 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 $S \in \mathcal{L}(V)$ 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 7.91 and then using the formula for T^* given by 7.74 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 7.91. ■

Exercise 17 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

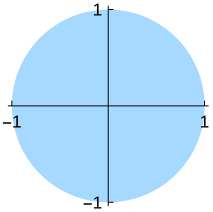
7.92 definition: *ball, B*

The *ball* in V centered at 0 with radius 1, denoted 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 $\mathbf{R}^2$.

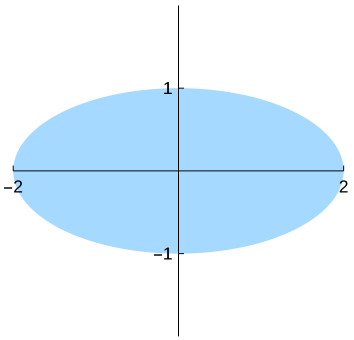
7.93 definition: *ellipsoid, $E(s_1 f_1, \dots, s_n f_n)$, principal axes*

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

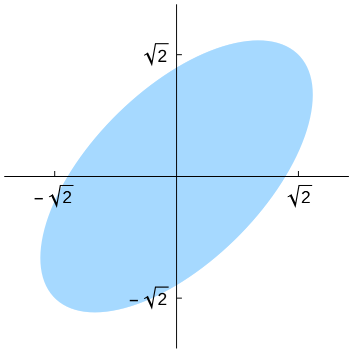
$$E(s_1 f_1, \dots, s_n f_n) = \left\{v \in V : \frac{|\langle v, f_1\rangle|^2}{s_1^2} + \cdots + \frac{|\langle v, f_n\rangle|^2}{s_n^2} < 1\right\}.$$

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 upon V. However, the inner product space V should be clear from the context and also from the requirement that $f_1, \dots, f_n$ is an orthonormal basis of V.

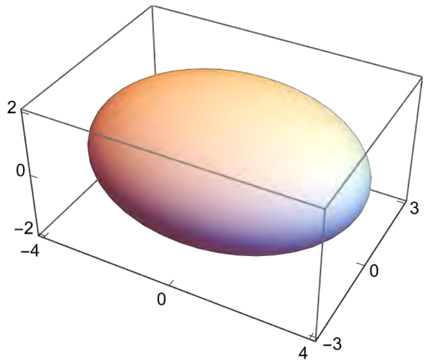
7.94 example: *ellipsoids*



The ellipsoid $E(2f_1, f_2)$ in $\mathbf{R}^2$, where f_1, f_2 is the standard basis of $\mathbf{R}^2$.



The ellipsoid $E(2f_1, f_2)$ in $\mathbf{R}^2$, where $f_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and $f_2 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.



The ellipsoid $E(4f_1, 3f_2, 2f_3)$ in $\mathbf{R}^3$, where f_1, f_2, f_3 is the standard basis of $\mathbf{R}^3$.

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 6.30(b)].

7.95 notation: $T(\Omega)$

For T a function defined on V and $\Omega \subset V$, define $T(\Omega)$ by

$$T(\Omega) = \{Tv : v \in \Omega\}.$$

Thus if T is a function defined on V, then $T(V) = \operatorname{range} T$.

The next result states that each invertible operator $T \in \mathcal{L}(V)$ maps the ball B in V onto an ellipsoid in V, with the principal axes of this ellipsoid coming from the singular value decomposition of T.

7.96 *invertible operator takes ball to ellipsoid*

Suppose $T \in \mathcal{L}(V)$ is invertible. Then T maps the ball B in V onto an ellipsoid in V.

Proof Suppose T has singular value decomposition

7.97 $$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n$$

for all $v \in V$, where $s_1, \dots, s_n$ are the singular values of T and $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are both orthonormal bases of V. We will show that $T(B) = E(s_1 f_1, \dots, s_n f_n)$.

First suppose $v \in B$. Then

$$\frac{|\langle Tv, f_1\rangle|^2}{s_1^2} + \cdots + \frac{|\langle Tv, f_n\rangle|^2}{s_n^2} = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 < 1.$$

Thus $Tv \in E(s_1 f_1, \dots, s_n f_n)$, and hence $T(B) \subset E(s_1 f_1, \dots, s_n f_n)$.

To prove inclusion in the other direction, now suppose $w \in E(s_1 f_1, \dots, s_n f_n)$. 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 7.97 implies that $Tv = \langle w, f_1\rangle f_1 + \cdots + \langle w, f_n\rangle f_n = w$. Thus $T(B) \supset E(s_1 f_1, \dots, s_n f_n)$. ∎

We now use the previous result to show that invertible operators take all ellipsoids, not just the ball of radius 1, to ellipsoids.

7.98 *invertible operator takes ellipsoids to ellipsoids*

Suppose $T \in \mathcal{L}(V)$ is invertible and E is an ellipsoid in V. Then $T(E)$ is an ellipsoid in V.

Proof There is 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 $S \in \mathcal{L}(V)$ 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 is easy to verify. Thus

$$T(E) = T(S(B)) = (TS)(B).$$

The equation above and 7.96, applied to TS, show that $T(E)$ is an ellipsoid in V. ■

Recall (see 3.75) that if $v \in V$ and $\Omega \subset V$ then $v + \Omega$ is defined by

$$v + \Omega = \{v + u : u \in \Omega\}.$$

Geometrically, the sets Ω and $v + \Omega$ 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*.

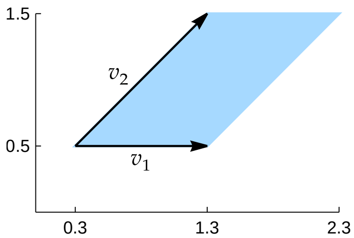
7.99 definition: $P(v_1, \ldots, v_n)$, *parallelepiped*

Suppose $v_1, \ldots, v_n$ is a basis of V. Let

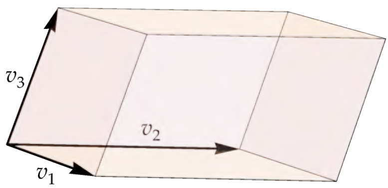
$$P(v_1, \ldots, v_n) = \{a_1 v_1 + \cdots + a_n v_n : a_1, \ldots, a_n \in (0, 1)\}.$$

A *parallelepiped* is a set of the form $v + P(v_1, \ldots, v_n)$ for some $v \in V$. The vectors $v_1, \ldots, v_n$ are called the *edges* of this parallelepiped.

7.100 example: *parallelepipeds*



The parallelepiped $(0.3, 0.5) + P((1,0), (1,1))$ *in* $\mathbf{R}^2$.



A parallelepiped in $\mathbf{R}^3$.

7.101 *invertible operator takes parallelepipeds to parallelepipeds*

Suppose $v \in V$ and $v_1, \ldots, v_n$ is a basis of V. Suppose $T \in \mathcal{L}(V)$ is invertible. Then

$$T(v + P(v_1, \ldots, v_n)) = Tv + P(Tv_1, \ldots, Tv_n).$$

Proof Because T is invertible, the list $Tv_1, \ldots, Tv_n$ is a basis of V. The linearity of T implies that

$$T(v + a_1 v_1 + \cdots + a_n v_n) = Tv + a_1 Tv_1 + \cdots + a_n Tv_n$$

for all $a_1, \ldots, a_n \in (0, 1)$. Thus $T(v + P(v_1, \ldots, v_n)) = Tv + P(Tv_1, \ldots, Tv_n)$. ■

Just as the rectangles are distinguished among the parallelograms in $\mathbf{R}^2$, we give a special name to the parallelepipeds in V whose defining edges are orthogonal to each other.

7.102 definition: *box*

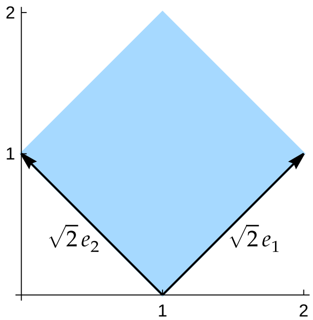
A *box* in V is a set of the form

$$v + P(r_1 e_1, \ldots, r_n e_n),$$

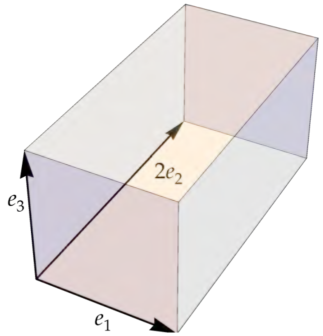
where $v \in V$ and $r_1, \ldots, r_n$ are positive numbers and $e_1, \ldots, e_n$ is an orthonormal basis of V.

Note that in the special case of $\mathbf{R}^2$ each box is a rectangle, but the terminology *box* can be used in all dimensions.

7.103 example: *boxes*



The box $(1,0) + P(\sqrt{2}\, e_1, \sqrt{2}\, e_2)$, *where* $e_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ *and* $e_2 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.



The box $P(e_1, 2e_2, e_3)$, *where* e_1, e_2, e_3 *is the standard basis of* $\mathbf{R}^3$.

Suppose $T \in \mathcal{L}(V)$ is invertible. Then T maps every parallelepiped in V to a parallelepiped in V (by 7.101). 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.

7.104 *each operator takes some boxes to boxes*

Suppose $T \in \mathcal{L}(V)$ 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 $s_1, \ldots, s_n$ are the singular values of T and $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ are orthonormal bases of V and the equation above holds for all $v \in V$. Then T maps the box $v + P(r_1e_1, \ldots, r_ne_n)$ onto the box $Tv + P(r_1s_1f_1, \ldots, r_ns_nf_n)$ for all positive numbers $r_1, \ldots, r_n$ and all $v \in V$.

Proof If $a_1, \ldots, a_n \in (0, 1)$ and $r_1, \ldots, r_n$ are positive numbers and $v \in V$, then

$$T(v + a_1r_1e_1 + \cdots + a_nr_ne_n) = Tv + a_1r_1s_1f_1 + \cdots + a_nr_ns_nf_n.$$

Thus $T(v + P(r_1e_1, \ldots, r_ne_n)) = Tv + P(r_1s_1f_1, \ldots, r_ns_nf_n)$. ■

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 $\mathbf{F} = \mathbf{R}$ 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 $\mathbf{R}^3$. 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.

7.105 definition: *volume of a box*

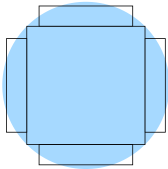
Suppose $\mathbf{F} = \mathbf{R}$. If $v \in V$ and $r_1, \ldots, r_n$ are positive numbers and $e_1, \ldots, e_n$ is an orthonormal basis of V, then

$$\text{volume}(v + P(r_1e_1, \ldots, r_ne_n)) = r_1 \times \cdots \times r_n.$$

The definition above agrees with the familiar formulas for the area (which we are calling the volume) of a rectangle in $\mathbf{R}^2$ and for the volume of a box in $\mathbf{R}^3$. For example, the first box in Example 7.103 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 7.103 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 of this ball ≈ sum of the volumes of the five boxes.

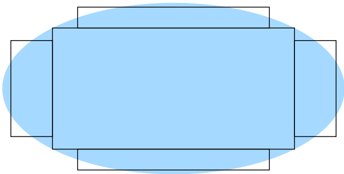
7.106 definition: *volume*

Suppose $\mathbf{F} = \mathbf{R}$ and $\Omega \subset V$. Then the *volume* of Ω, denoted $\operatorname{volume}\Omega$, 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 some 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 $\operatorname{volume}(\Omega_1 \cup \Omega_1) = \operatorname{volume}\Omega_1 + \operatorname{volume}\Omega_2$? 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.

7.107 example: *volume change by a linear map*

Suppose that $T \in \mathcal{L}(\mathbf{R}^2)$ 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 $\mathbf{R}^2$. This linear map stretches by a factor of 2 along the e_1 axis. 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 with twice the width and the same height as in the original figure. Hence each box gets mapped to a box with 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.



Each box here has twice the width and the same height as the boxes in the previous figure.

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 7.104. This observation leads to a natural proof of the following result.

7.108 *volume changes by a factor of the product of the singular values*

Suppose $\mathbf{F} = \mathbf{R}$, $T \in \mathcal{L}(V)$ is invertible, and $\Omega \subset V$. 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 $v + P(r_1e_1, \ldots, r_ne_n)$, which have volume $r_1 \times \cdots \times r_n$. The operator T maps each box $v + P(r_1e_1, \ldots, r_ne_n)$ onto the box $Tv + \mathcal{P}(r_1s_1f_1, \ldots, r_ns_nf_n)$, 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. ■

Exercises 7F

1 Prove that if $S, T \in \mathcal{L}(V, W)$, then $\big|\|S\| - \|T\|\big| \le \|S - T\|$.

The inequality above is called the ***reverse triangle inequality****.*

2 Suppose that $T \in \mathcal{L}(V)$ is self-adjoint or that $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that

$$\|T\| = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } T\}.$$

3 Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$ with $v \neq 0$. Prove that $\|Tv\| = \|T\|\,\|v\|$ if and only if v is an eigenvector of T^*T corresponding to the largest eigenvalue of T^*T.

4 Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $T \in \mathcal{L}(V, W)$. Prove that

$$\max\{\|Te_1\|, \ldots, \|Te_n\|\} \le \|T\| \le \left(\|Te_1\|^2 + \cdots + \|Te_n\|^2\right)^{1/2}.$$

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 11 in Section 7E.

5 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 $\mathbf{F}$ as an inner product space with $\langle \alpha, \beta\rangle = \alpha\overline{\beta}$ for all $\alpha, \beta \in \mathbf{F}$. Thus $\|\varphi_u\|$ means the norm of φ_u as a linear map in $\mathcal{L}(V, \mathbf{F})$.

6 Suppose U is an inner product space, $T \in \mathcal{L}(V, U)$ and $S \in \mathcal{L}(U, W)$. Prove that

$$\|ST\| \leq \|S\|\,\|T\|.$$

7 Prove or give counterexample: If $S, T \in \mathcal{L}(V)$, then $\|ST\| = \|TS\|$.

8 Prove that if $T \in \mathcal{L}(V, W)$, then $\|T^*T\| = \|T\|^2$.

This formula for $\|T^*T\|$ leads to the important subject of C^*-algebras.

9 Suppose $T \in \mathcal{L}(V)$ is normal. Prove that $\|T^k\| = \|T\|^k$ for every positive integer k.

10 Suppose $T \in \mathcal{L}(V)$ is a positive operator. Show that $\left\|\sqrt{T}\right\| = \sqrt{\|T\|}$.

11 Suppose $S, T \in \mathcal{L}(V)$ are positive. Show that $\|S + T\| \geq \max\{\|S\|, \|T\|\}$.

12 Suppose $T \in \mathcal{L}(V)$ and $\epsilon > 0$. Prove that there exists an invertible operator $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$.

13 Suppose $\dim V > 1$. Suppose $T \in \mathcal{L}(V)$ is not invertible and $\epsilon > 0$. Prove that there exists $S \in \mathcal{L}(V)$ such that S is not invertible and $0 < \|T - S\| < \epsilon$.

14 Suppose $\mathbf{F} = \mathbf{C}$. Suppose $T \in \mathcal{L}(V)$ and $\epsilon > 0$. Prove that there exists a diagonalizable operator $S \in \mathcal{L}(V)$ such that $0 < \|T - S\| < \epsilon$.

15 Suppose $\dim V > 1$ and $\dim W > 1$. Prove that the norm on $\mathcal{L}(V, W)$ does not come from an inner product. In other words, prove that there does not exist an inner product on $\mathcal{L}(V, W)$ such that

$$\max\{\|Tv\| : v \in V \text{ and } \|v\| \leq 1\} = \sqrt{\langle T, T\rangle}$$

for all $T \in \mathcal{L}(V, W)$.

16 Suppose $T \in \mathcal{L}(V, W)$. Let $n = \dim V$ and let $s_1 \geq \cdots \geq s_n$ denote the singular values of T. Prove that if $1 \leq k \leq n$, then

$$\min\{\|T|_U\| : U \text{ is a subspace of } V \text{ with } \dim U = k\} = s_{n-k+1}.$$

17 Suppose $T \in \mathcal{L}(V)$ 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 $S \in \mathcal{L}(V)$ 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 $E \in \mathcal{L}(V)$ 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.

18 Suppose $T \in \mathcal{L}(V, W)$. 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 22 in Section 6B).

19 Suppose $V \neq \{0\}$ and $T \in \mathcal{L}(V)$ is invertible. Prove that

$$\|T^{-1}\| = \|T\|^{-1} \iff \frac{T}{\|T\|} \text{ is a unitary operator.}$$

20 Fix $u, x \in V$ with $u \neq 0$. Define $T \in \mathcal{L}(V)$ by $Tv = \langle v, u \rangle x$ for every $v \in V$. Prove that

$$\sqrt{T^*T}v = \frac{\|x\|}{\|u\|}\langle v, u \rangle u$$

for every $v \in V$.

21 Suppose $T \in \mathcal{L}(V)$. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $T = \sqrt{TT^*}\, S$.

22 Suppose $T \in \mathcal{L}(V)$. Prove that T is invertible if and only if there exists a unique unitary operator $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$.

23 Suppose $T \in \mathcal{L}(V)$.

(a) Use the polar decomposition to show that there exists a unitary operator $S \in \mathcal{L}(V)$ such that $TT^* = ST^*TS^*$.

(b) Show how part (a) implies that T and T^* have the same singular values.

24 Define $T \in \mathcal{L}(\mathbf{F}^3)$ by

$$T(z_1, z_2, z_3) = (z_3, 2z_1, 3z_2).$$

Find (explicitly) a unitary operator $S \in \mathcal{L}(\mathbf{F}^3)$ such that $T = S\sqrt{T^*T}$.

25 Suppose $T \in \mathcal{L}(V)$, $S \in \mathcal{L}(V)$ is a unitary operator, and $R \in \mathcal{L}(V)$ 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 7.90), then the positive operator equals $\sqrt{T^*T}$.

26 Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal. Prove that there exists a unitary operator $S \in \mathcal{L}(V)$ 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.

27 Suppose that $T \in \mathcal{L}(V, W)$ 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 $(\operatorname{null} T)^\perp$ such that

$$T\Bigl(E\bigl(\frac{e_1}{s_1}, \ldots, \frac{e_m}{s_m}\bigr)\Bigr)$$

equals the ball in range T centered at 0 with radius 1.