Linear Algebra Abridged

Sheldon Axler

This file is generated from *Linear Algebra Done Right* (third edition) by excluding all proofs, examples, and exercises, along with most comments. Learning linear algebra without proofs, examples, and exercises is probably impossible. Thus this abridged version should not substitute for the full book. However, this abridged version may be useful to students seeking to review the statements of the main results of linear algebra.

As a visual aid, definitions are in beige boxes and theorems are in blue boxes. The numbering of definitions and theorems is the same as in the full book. Thus 1.1 is followed in this abridged version by 1.3 (the missing 1.2 corresponds to an example in the full version that is not present here).

The full version of *Linear Algebra Done Right* is available at springer.com and amazon.com in both printed and electronic forms. A free sample chapter of the full version, and other information, is available at the book's website: http://linear.axler.net.

13 March 2016

Contents

Linear Algebra Abridged is generated from *Linear Algebra Done Right* (by Sheldon Axler, third edition) by excluding all proofs, examples, and exercises, along with most comments. The full version of *Linear Algebra Done Right* is available at springer.com and amazon.com in both printed and electronic forms.

CHAPTER

1

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

Vector Spaces

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

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

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

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

LEARNING OBJECTIVES FOR THIS CHAPTER

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

Linear Algebra Abridged is generated from *Linear Algebra Done Right* (by Sheldon Axler, third edition) by excluding all proofs, examples, and exercises, along with most comments. The full version of *Linear Algebra Done Right* is available at springer.com and amazon.com in both printed and electronic forms.

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

Complex Numbers

1.1 Definition ***complex numbers***

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

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

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

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

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

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

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

1.3 Properties of complex arithmetic

commutativity

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

associativity

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

identities

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

additive inverse

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

multiplicative inverse

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

distributive property

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

The properties above are proved using the familiar properties of real numbers and the definitions of complex addition and multiplication.

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

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

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

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

1.6 Notation $\mathbf{F}$

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

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

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

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

Lists

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

1.8 Definition ***list, length***

Suppose n is a nonnegative integer. A ***list*** of ***length*** n is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses. A list of length n looks like this:

$$(x_1, \dots, x_n).$$

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

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

Sometimes we will use the word ***list*** without specifying its length. Remember, however, that by definition each list has a finite length that is a nonnegative integer. Thus an object that looks like

$$(x_1, x_2, \dots),$$

which might be said to have infinite length, is not a list.

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

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

$\mathbf{F}^n$

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

1.10 Definition $\mathbf{F}^n$

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

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

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

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

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

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

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

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

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

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

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

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

1.14 Definition 0

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

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

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

A picture can aid our intuition. We will draw pictures in $\mathbf{R}^2$ because we can sketch this space on 2-dimensional surfaces such as paper and blackboards. A typical element of $\mathbf{R}^2$ is a point $x = (x_1, x_2)$. Sometimes we think of x

not as a point but as an arrow starting at the origin and ending at (x_1, x_2), as shown here. When we think of x as an arrow, we refer to it as a ***vector***.

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

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

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

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

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

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

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

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

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

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

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

Having dealt with addition in $\mathbf{F}^n$, we now turn to multiplication. We could define a multiplication in $\mathbf{F}^n$ in a similar fashion, starting with two elements of $\mathbf{F}^n$ and getting another element of $\mathbf{F}^n$ by multiplying corresponding coordinates. Experience shows that this definition is not useful for our purposes.

Another type of multiplication, called scalar multiplication, will be central to our subject. Specifically, we need to define what it means to multiply an element of $\mathbf{F}^n$ by an element of $\mathbf{F}$.

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

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

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

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

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

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

Digression on Fields

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

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

1.B *Definition of Vector Space*

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

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

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

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

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

1.19 Definition ***vector space***

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

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

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

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

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

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

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

The following geometric language sometimes aids our intuition.

1.20 Definition ***vector, point***

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

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

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

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

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

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

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

1.23 Notation $\mathbf{F}^S$

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

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

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

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

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

1.25 Unique additive identity

A vector space has a unique additive identity.

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

1.26 Unique additive inverse

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

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

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

Let $v, w \in V$. Then

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

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

1.28 Notation V

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

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

1.29 The number 0 times a vector

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

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

1.30 A number times the vector 0

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

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

1.31 The number -1 times a vector

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

1.C *Subspaces*

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

1.32 Definition ***subspace***

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

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

1.34 Conditions for a subspace

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

additive identity
$0 \in U$;

closed under addition
$u, w \in U$ implies $u + w \in U$;

closed under scalar multiplication
$a \in \mathbf{F}$ and $u \in U$ implies $au \in U$.

The three conditions in the result above usually enable us to determine quickly whether a given subset of V is a subspace of V.

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

Sums of Subspaces

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

1.36 Definition *sum of subsets*

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

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

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

1.39 Sum of subspaces is the smallest containing subspace

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

Direct Sums

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

$$u_1 + \cdots + u_m,$$

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

1.40 Definition ***direct sum***

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

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

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

1.44 Condition for a direct sum

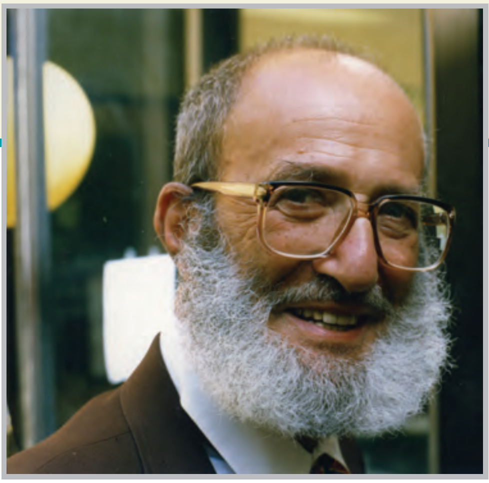
Suppose $U_1, \ldots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each u_j is in U_j, is by taking each u_j equal to 0.

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

1.45 Direct sum of two subspaces

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

CHAPTER
2



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

Finite-Dimensional Vector Spaces

Let's review our standing assumptions:

2.1 Notation $\mathbf{F}$, V

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

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

LEARNING OBJECTIVES FOR THIS CHAPTER

- span
- linear independence
- bases
- dimension

2.A *Span and Linear Independence*

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

2.2 Notation ***list of vectors***

We will usually write lists of vectors without surrounding parentheses.

Linear Combinations and Span

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

2.3 Definition ***linear combination***

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

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

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

2.5 Definition ***span***

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

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

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

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

2.7 Span is the smallest containing subspace

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

2.8 Definition ***spans***

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Linear Independence

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

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

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

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

Subtracting the last two equations, we have

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

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

2.17 Definition ***linearly independent***

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

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

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

2.19 Definition ***linearly dependent***

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

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

2.21 Linear Dependence Lemma

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

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

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

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

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

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

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

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

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

2.26 Finite-dimensional subspaces

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

2.B *Bases*

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

2.27 Definition ***basis***

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

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

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

2.29 Criterion for basis

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

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

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

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

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

2.31 Spanning list contains a basis

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

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

2.32 Basis of finite-dimensional vector space

Every finite-dimensional vector space has a basis.

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

2.33 Linearly independent list extends to a basis

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

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

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

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

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

2.C Dimension

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

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

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

2.35 Basis length does not depend on basis

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

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

2.36 Definition ***dimension,*** $\dim V$

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

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

2.38 Dimension of a subspace

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

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

2.39 Linearly independent list of the right length is a basis

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

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

2.42 Spanning list of the right length is a basis

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

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

2.43 Dimension of a sum

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

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

CHAPTER

3

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

Linear Maps

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

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

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

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

LEARNING OBJECTIVES FOR THIS CHAPTER

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

3.A The Vector Space of Linear Maps

Definition and Examples of Linear Maps

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

3.2 Definition ***linear map***

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

additivity

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

homogeneity

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

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

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

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

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

3.5 Linear maps and basis of domain

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

$$Tv_j = w_j$$

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

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

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

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

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

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

for all $v \in V$.

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

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

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

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

Note that the additive identity of $\mathcal{L}(V, W)$ is the zero linear map defined earlier in this section.

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

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

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

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

for $u \in U$.

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

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

3.9 Algebraic properties of products of linear maps

associativity

$$(T_1T_2)T_3 = T_1(T_2T_3)$$

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

identity

$$TI = IT = T$$

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

distributive properties

$$(S_1 + S_2)T = S_1T + S_2T \quad \text{and} \quad S(T_1 + T_2) = ST_1 + ST_2$$

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

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

3.11 Linear maps take 0 to 0

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

3.B *Null Spaces and Ranges*

Null Space and Injectivity

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

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

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

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

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

3.14 The null space is a subspace

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

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

3.15 Definition ***injective***

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

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

The next result says that we can check whether a linear map is injective by checking whether 0 is the only vector that gets mapped to 0.

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

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

Range and Surjectivity

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

3.17 Definition ***range***

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

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

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

3.19 The range is a subspace

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

3.20 Definition ***surjective***

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

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

Fundamental Theorem of Linear Maps

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

3.22 Fundamental Theorem of Linear Maps

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

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

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

3.23 A map to a smaller dimensional space is not injective

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

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

3.24 A map to a larger dimensional space is not surjective

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

As we will now see, 3.23 and 3.24 have important consequences in the theory of linear equations. The idea here is to express questions about systems of linear equations in terms of linear maps.

3.26 Homogeneous system of linear equations

A homogeneous system of linear equations with more variables than equations has nonzero solutions.

Example of the result above: a homogeneous system of four linear equations with five variables has nonzero solutions.

3.29 Inhomogeneous system of linear equations

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

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

3.C *Matrices*

Representing a Linear Map by a Matrix

We know that if $v_1, \dots, v_n$ is a basis of V and $T : V \to W$ is linear, then the values of $Tv_1, \dots, Tv_n$ determine the values of T on arbitrary vectors in V (see 3.5). As we will soon see, matrices are used as an efficient method of recording the values of the Tv_j's in terms of a basis of W.

3.30 Definition ***matrix,*** $A_{j,k}$

Let m and n denote positive integers. An m-by-n ***matrix*** A is a rectangular array of elements of $\mathbf{F}$ with m rows and n columns:

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix}.$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A. In other words, the first index refers to the row number and the second index refers to the column number.

Thus $A_{2,3}$ refers to the entry in the second row, third column of a matrix A.

Now we come to the key definition in this section.

3.32 Definition ***matrix of a linear map,*** $\mathcal{M}(T)$

Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. The ***matrix of*** T with respect to these bases is the m-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases are not clear from the context, then the notation $\mathcal{M}\big(T, (v_1, \dots, v_n), (w_1, \dots, w_m)\big)$ is used.

The matrix $\mathcal{M}(T)$ of a linear map $T \in \mathcal{L}(V, W)$ depends on the basis $v_1, \dots, v_n$ of V and the basis $w_1, \dots, w_m$ of W, as well as on T. However, the bases should be clear from the context, and thus they are often not included in the notation.

To remember how $\mathcal{M}(T)$ is constructed from T, you might write across the top of the matrix the basis vectors $v_1, \dots, v_n$ for the domain and along the left the basis vectors $w_1, \dots, w_m$ for the vector space into which T maps, as follows:

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & \dots & v_k & \dots & v_n \end{array} \\ \left(\begin{array}{ccccc} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{array} \right) \end{array}.$$

In the matrix above only the k^{th} column is shown. Thus the second index of each displayed entry of the matrix above is k. The picture above should remind you that Tv_k can be computed from $\mathcal{M}(T)$ by multiplying each entry in the k^{th} column by the corresponding w_j from the left column, and then adding up the resulting vectors.

If T is a linear map from $\mathbf{F}^n$ to $\mathbf{F}^m$, then unless stated otherwise, assume the bases in question are the standard ones (where the k^{th} basis vector is 1 in the k^{th} slot and 0 in all the other slots). If you think of elements of $\mathbf{F}^m$ as columns of m numbers, then you can think of the k^{th} column of $\mathcal{M}(T)$ as T applied to the k^{th} standard basis vector.

When working with $\mathcal{P}_m(\mathbf{F})$, use the standard basis $1, x, x^2, \dots, x^m$ unless the context indicates otherwise.

Addition and Scalar Multiplication of Matrices

For the rest of this section, assume that V and W are finite-dimensional and that a basis has been chosen for each of these vector spaces. Thus for each linear map from V to W, we can talk about its matrix (with respect to the chosen bases, of course). Is the matrix of the sum of two linear maps equal to the sum of the matrices of the two maps?

Right now this question does not make sense, because although we have defined the sum of two linear maps, we have not defined the sum of two matrices. Fortunately, the obvious definition of the sum of two matrices has the right properties. Specifically, we make the following definition.

3.35 Definition ***matrix addition***

The ***sum of two matrices of the same size*** is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \dots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \dots & C_{m,n} \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1} + C_{1,1} & \dots & A_{1,n} + C_{1,n} \\ \vdots & & \vdots \\ A_{m,1} + C_{m,1} & \dots & A_{m,n} + C_{m,n} \end{pmatrix}.$$

In other words, $(A + C)_{j,k} = A_{j,k} + C_{j,k}$.

In the following result, the assumption is that the same bases are used for all three linear maps $S + T$, S, and T.

3.36 The matrix of the sum of linear maps

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

Still assuming that we have some bases in mind, is the matrix of a scalar times a linear map equal to the scalar times the matrix of the linear map? Again the question does not make sense, because we have not defined scalar multiplication on matrices. Fortunately, the obvious definition again has the right properties.

3.37 Definition ***scalar multiplication of a matrix***

The product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \dots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \dots & \lambda A_{m,n} \end{pmatrix}.$$

In other words, $(\lambda A)_{j,k} = \lambda A_{j,k}$.

In the following result, the assumption is that the same bases are used for both linear maps λT and T.

3.38 The matrix of a scalar times a linear map

Suppose $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.

Because addition and scalar multiplication have now been defined for matrices, you should not be surprised that a vector space is about to appear. We need only a bit of notation so that this new vector space has a name.

3.39 Notation $\mathbf{F}^{m,n}$

For m and n positive integers, the set of all m-by-n matrices with entries in $\mathbf{F}$ is denoted by $\mathbf{F}^{m,n}$.

3.40 $\dim \mathbf{F}^{m,n} = mn$

Suppose m and n are positive integers. With addition and scalar multiplication defined as above, $\mathbf{F}^{m,n}$ is a vector space with dimension mn.

Matrix Multiplication

Suppose, as previously, that $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. Suppose also that we have another vector space U and that $u_1, \dots, u_p$ is a basis of U.

Consider linear maps $T : U \to V$ and $S : V \to W$. The composition ST is a linear map from U to W. Does $\mathcal{M}(ST)$ equal $\mathcal{M}(S)\mathcal{M}(T)$? This question does not yet make sense, because we have not defined the product of two matrices. We will choose a definition of matrix multiplication that forces this question to have a positive answer. Let's see how to do this.

Suppose $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$. For $1 \le k \le p$, we have

$$\begin{aligned}(ST)u_k &= S\Big(\sum_{r=1}^{n} C_{r,k} v_r\Big) \\ &= \sum_{r=1}^{n} C_{r,k} S v_r \\ &= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r} w_j \\ &= \sum_{j=1}^{m} \Big(\sum_{r=1}^{n} A_{j,r} C_{r,k}\Big) w_j.\end{aligned}$$

Thus $\mathcal{M}(ST)$ is the m-by-p matrix whose entry in row j, column k, equals

$$\sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

Now we see how to define matrix multiplication so that the desired equation $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$ holds.

3.41 Definition ***matrix multiplication***

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then AC is defined to be the m-by-p matrix whose entry in row j, column k, is given by the following equation:

$$(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}.$$

In other words, the entry in row j, column k, of AC is computed by taking row j of A and column k of C, multiplying together corresponding entries, and then summing.

Note that we define the product of two matrices only when the number of columns of the first matrix equals the number of rows of the second matrix.

Matrix multiplication is not commutative. In other words, AC is not necessarily equal to CA even if both products are defined. Matrix multiplication is distributive and associative.

In the following result, the assumption is that the same basis of V is used in considering $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, the same basis of W is used in considering $S \in \mathcal{L}(V, W)$ and $ST \in \mathcal{L}(U, W)$, and the same basis of U is used in considering $T \in \mathcal{L}(U, V)$ and $ST \in \mathcal{L}(U, W)$.

3.43 The matrix of the product of linear maps

If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.

The proof of the result above is the calculation that was done as motivation before the definition of matrix multiplication.

In the next piece of notation, note that as usual the first index refers to a row and the second index refers to a column, with a vertically centered dot used as a placeholder.

3.44 Notation $A_{j,\cdot}$, $A_{\cdot,k}$

Suppose A is an m-by-n matrix.

- If $1 \leq j \leq m$, then $A_{j,\cdot}$ denotes the 1-by-n matrix consisting of row j of A.
- If $1 \leq k \leq n$, then $A_{\cdot,k}$ denotes the m-by-1 matrix consisting of column k of A.

The product of a 1-by-n matrix and an n-by-1 matrix is a 1-by-1 matrix. However, we will frequently identify a 1-by-1 matrix with its entry.

Our next result gives another way to think of matrix multiplication: the entry in row j, column k, of AC equals (row j of A) times (column k of C).

3.47 Entry of matrix product equals row times column

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{j,k} = A_{j,\cdot} C_{\cdot,k}$$

for $1 \leq j \leq m$ and $1 \leq k \leq p$.

The next result gives yet another way to think of matrix multiplication. It states that column k of AC equals A times column k of C.

3.49 Column of matrix product equals matrix times column

Suppose A is an m-by-n matrix and C is an n-by-p matrix. Then

$$(AC)_{\cdot,k} = AC_{\cdot,k}$$

for $1 \leq k \leq p$.

We give one more way of thinking about the product of an m-by-n matrix and an n-by-1 matrix.

3.52 Linear combination of columns

Suppose A is an m-by-n matrix and $c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$ is an n-by-1 matrix. Then

$$Ac = c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}.$$

In other words, Ac is a linear combination of the columns of A, with the scalars that multiply the columns coming from c.

3.D *Invertibility and Isomorphic Vector Spaces*

Invertible Linear Maps

We begin this section by defining the notions of invertible and inverse in the context of linear maps.

3.53 Definition ***invertible, inverse***

- A linear map $T \in \mathcal{L}(V, W)$ is called ***invertible*** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that ST equals the identity map on V and TS equals the identity map on W.
- A linear map $S \in \mathcal{L}(W, V)$ satisfying $ST = I$ and $TS = I$ is called an ***inverse*** of T (note that the first I is the identity map on V and the second I is the identity map on W).

3.54 Inverse is unique

An invertible linear map has a unique inverse.

Now that we know that the inverse is unique, we can give it a notation.

3.55 Notation T^{-1}

If T is invertible, then its inverse is denoted by T^{-1}. In other words, if $T \in \mathcal{L}(V, W)$ is invertible, then T^{-1} is the unique element of $\mathcal{L}(W, V)$ such that $T^{-1}T = I$ and $TT^{-1} = I$.

The following result characterizes the invertible linear maps.

3.56 Invertibility is equivalent to injectivity and surjectivity

A linear map is invertible if and only if it is injective and surjective.

Isomorphic Vector Spaces

The next definition captures the idea of two vector spaces that are essentially the same, except for the names of the elements of the vector spaces.

3.58 Definition ***isomorphism, isomorphic***

- An ***isomorphism*** is an invertible linear map.
- Two vector spaces are called ***isomorphic*** if there is an isomorphism from one vector space onto the other one.

Think of an isomorphism $T : V \to W$ as relabeling $v \in V$ as $Tv \in W$. This viewpoint explains why two isomorphic vector spaces have the same vector space properties. The terms "isomorphism" and "invertible linear map" mean the same thing. Use "isomorphism" when you want to emphasize that the two spaces are essentially the same.

3.59 Dimension shows whether vector spaces are isomorphic

Two finite-dimensional vector spaces over $\mathbf{F}$ are isomorphic if and only if they have the same dimension.

The previous result implies that each finite-dimensional vector space V is isomorphic to $\mathbf{F}^n$, where $n = \dim V$.

If $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W, then for each $T \in \mathcal{L}(V, W)$, we have a matrix $\mathcal{M}(T) \in \mathbf{F}^{m,n}$. In other words, once bases have been fixed for V and W, $\mathcal{M}$ becomes a function from $\mathcal{L}(V, W)$ to $\mathbf{F}^{m,n}$. Notice that 3.36 and 3.38 show that $\mathcal{M}$ is a linear map. This linear map is actually invertible, as we now show.

3.60 $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$ are isomorphic

Suppose $v_1, \ldots, v_n$ is a basis of V and $w_1, \ldots, w_m$ is a basis of W. Then $\mathcal{M}$ is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbf{F}^{m,n}$.

Now we can determine the dimension of the vector space of linear maps from one finite-dimensional vector space to another.

3.61 $\dim \mathcal{L}(V, W) = (\dim V)(\dim W)$

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

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Linear Maps Thought of as Matrix Multiplication

Previously we defined the matrix of a linear map. Now we define the matrix of a vector.

3.62 Definition ***matrix of a vector, $\mathcal{M}(v)$***

Suppose $v \in V$ and $v_1, \ldots, v_n$ is a basis of V. The ***matrix of*** v with respect to this basis is the n-by-1 matrix

$$\mathcal{M}(v) = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

where $c_1, \ldots, c_n$ are the scalars such that

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

The matrix $\mathcal{M}(v)$ of a vector $v \in V$ depends on the basis $v_1, \ldots, v_n$ of V, as well as on v. However, the basis should be clear from the context and thus it is not included in the notation.

Occasionally we want to think of elements of V as relabeled to be n-by-1 matrices. Once a basis $v_1, \dots, v_n$ is chosen, the function $\mathcal{M}$ that takes $v \in V$ to $\mathcal{M}(v)$ is an isomorphism of V onto $\mathbf{F}^{n,1}$ that implements this relabeling.

Recall that if A is an m-by-n matrix, then $A_{\cdot,k}$ denotes the k^{th} column of A, thought of as an m-by-1 matrix. In the next result, $\mathcal{M}(Tv_k)$ is computed with respect to the basis $w_1, \dots, w_m$ of W.

3.64 $\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(Tv_k)$.

Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. Let $1 \le k \le n$. Then the k^{th} column of $\mathcal{M}(T)$, which is denoted by $\mathcal{M}(T)_{\cdot,k}$, equals $\mathcal{M}(Tv_k)$.

The next result shows how the notions of the matrix of a linear map, the matrix of a vector, and matrix multiplication fit together.

3.65 Linear maps act like matrix multiplication

Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Suppose $v_1, \dots, v_n$ is a basis of V and $w_1, \dots, w_m$ is a basis of W. Then

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

Each m-by-n matrix A induces a linear map from $\mathbf{F}^{n,1}$ to $\mathbf{F}^{m,1}$, namely the matrix multiplication function that takes $x \in \mathbf{F}^{n,1}$ to $Ax \in \mathbf{F}^{m,1}$. The result above can be used to think of every linear map (from one finite-dimensional vector space to another finite-dimensional vector space) as a matrix multiplication map after suitable relabeling via the isomorphisms given by $\mathcal{M}$. Specifically, if $T \in \mathcal{L}(V, W)$ and we identify $v \in V$ with $\mathcal{M}(v) \in \mathbf{F}^{n,1}$, then the result above says that we can identify Tv with $\mathcal{M}(T)\mathcal{M}(v)$.

Because the result above allows us to think (via isomorphisms) of each linear map as multiplication on $\mathbf{F}^{n,1}$ by some matrix A, keep in mind that the specific matrix A depends not only on the linear map but also on the choice of bases. One of the themes of many of the most important results in later chapters will be the choice of a basis that makes the matrix A as simple as possible.

In this book, we concentrate on linear maps rather than on matrices. However, sometimes thinking of linear maps as matrices (or thinking of matrices as linear maps) gives important insights that we will find useful.

Operators

Linear maps from a vector space to itself are so important that they get a special name and special notation.

3.67 Definition ***operator,*** $\mathcal{L}(V)$

- A linear map from a vector space to itself is called an ***operator***.
- The notation $\mathcal{L}(V)$ denotes the set of all operators on V. In other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

A linear map is invertible if it is injective and surjective. For an operator, you might wonder whether injectivity alone, or surjectivity alone, is enough to imply invertibility. On infinite-dimensional vector spaces, neither condition alone implies invertibility.

In view of the example above, the next result is remarkable—it states that for operators on a finite-dimensional vector space, either injectivity or surjectivity alone implies the other condition. Often it is easier to check that an operator on a finite-dimensional vector space is injective, and then we get surjectivity for free.

3.69 Injectivity is equivalent to surjectivity in finite dimensions

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is invertible;

(b) T is injective;

(c) T is surjective.

3.E *Products and Quotients of Vector Spaces*

Products of Vector Spaces

As usual when dealing with more than one vector space, all the vector spaces in use should be over the same field.

3.71 Definition ***product of vector spaces***

Suppose $V_1, \dots, V_m$ are vector spaces over $\mathbf{F}$.

- The ***product*** $V_1 \times \cdots \times V_m$ is defined by

$$V_1 \times \cdots \times V_m = \{(v_1, \dots, v_m) : v_1 \in V_1, \dots, v_m \in V_m\}.$$

- Addition on $V_1 \times \cdots \times V_m$ is defined by

$$(u_1, \dots, u_m) + (v_1, \dots, v_m) = (u_1 + v_1, \dots, u_m + v_m).$$

- Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by

$$\lambda(v_1, \dots, v_m) = (\lambda v_1, \dots, \lambda v_m).$$

The next result should be interpreted to mean that the product of vector spaces is a vector space with the operations of addition and scalar multiplication as defined above.

3.73 Product of vector spaces is a vector space

Suppose $V_1, \dots, V_m$ are vector spaces over $\mathbf{F}$. Then $V_1 \times \cdots \times V_m$ is a vector space over $\mathbf{F}$.

Note that the additive identity of $V_1 \times \cdots \times V_m$ is $(0, \dots, 0)$, where the 0 in the j^{th} slot is the additive identity of V_j. The additive inverse of $(v_1, \dots, v_m) \in V_1 \times \cdots \times V_m$ is $(-v_1, \dots, -v_m)$.

3.76 Dimension of a product is the sum of dimensions

Suppose $V_1, \dots, V_m$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and

$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m.$$

Products and Direct Sums

In the next result, the map Γ is surjective by the definition of $U_1 + \cdots + U_m$. Thus the last word in the result below could be changed from “injective” to “invertible”.

3.77 Products and direct sums

Suppose that $U_1, \dots, U_m$ are subspaces of V. Define a linear map $\Gamma : U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by

$$\Gamma(u_1, \dots, u_m) = u_1 + \cdots + u_m.$$

Then $U_1 + \cdots + U_m$ is a direct sum if and only if Γ is injective.

3.78 A sum is a direct sum if and only if dimensions add up

Suppose V is finite-dimensional and $U_1, \dots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if

$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m.$$

In the special case $m = 2$, an alternative proof that $U_1 + U_2$ is a direct sum if and only if $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$ can be obtained by combining 1.45 and 2.43.

Quotients of Vector Spaces

We begin our approach to quotient spaces by defining the sum of a vector and a subspace.

3.79 Definition $v + U$

Suppose $v \in V$ and U is a subspace of V. Then $v + U$ is the subset of V defined by

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

3.81 Definition ***affine subset, parallel***

- An ***affine subset*** of V is a subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V.
- For $v \in V$ and U a subspace of V, the affine subset $v + U$ is said to be ***parallel*** to U.

3.83 Definition ***quotient space,*** V/U

Suppose U is a subspace of V. Then the ***quotient space*** V/U is the set of all affine subsets of V parallel to U. In other words,

$$V/U = \{v + U : v \in V\}.$$

Our next goal is to make V/U into a vector space. To do this, we will need the following result.

3.85 Two affine subsets parallel to U are equal or disjoint

Suppose U is a subspace of V and $v, w \in V$. Then the following are equivalent:

(a) $v - w \in U$;

(b) $v + U = w + U$;

(c) $(v + U) \cap (w + U) \neq \varnothing$.

Now we can define addition and scalar multiplication on V/U.

3.86 Definition ***addition and scalar multiplication on*** V/U

Suppose U is a subspace of V. Then ***addition*** and ***scalar multiplication*** are defined on V/U by

$$(v + U) + (w + U) = (v + w) + U$$
$$\lambda(v + U) = (\lambda v) + U$$

for $v, w \in V$ and $\lambda \in \mathbf{F}$.

As part of the proof of the next result, we will show that the definitions above make sense.

3.87 Quotient space is a vector space

Suppose U is a subspace of V. Then V/U, with the operations of addition and scalar multiplication as defined above, is a vector space.

The next concept will give us an easy way to compute the dimension of V/U.

3.88 Definition ***quotient map, π***

Suppose U is a subspace of V. The ***quotient map*** π is the linear map $\pi : V \to V/U$ defined by

$$\pi(v) = v + U$$

for $v \in V$.

The reader should verify that π is indeed a linear map. Although π depends on U as well as V, these spaces are left out of the notation because they should be clear from the context.

3.89 Dimension of a quotient space

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim V/U = \dim V - \dim U.$$

Each linear map T on V induces a linear map $\tilde{T}$ on $V/(\operatorname{null} T)$, which we now define.

3.90 Definition $\tilde{T}$

Suppose $T \in \mathcal{L}(V, W)$. Define $\tilde{T} : V/(\operatorname{null} T) \to W$ by

$$\tilde{T}(v + \operatorname{null} T) = Tv.$$

To show that the definition of $\tilde{T}$ makes sense, suppose $u, v \in V$ are such that $u + \operatorname{null} T = v + \operatorname{null} T$. By 3.85, we have $u - v \in \operatorname{null} T$. Thus $T(u - v) = 0$. Hence $Tu = Tv$. Thus the definition of $\tilde{T}$ indeed makes sense.

3.91 Null space and range of $\tilde{T}$

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

(a) $\tilde{T}$ is a linear map from $V/(\operatorname{null} T)$ to W;

(b) $\tilde{T}$ is injective;

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

(d) $V/(\operatorname{null} T)$ is isomorphic to $\operatorname{range} T$.

3.F Duality

The Dual Space and the Dual Map

Linear maps into the scalar field $\mathbf{F}$ play a special role in linear algebra, and thus they get a special name:

3.92 Definition ***linear functional***

A ***linear functional*** on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

The vector space $\mathcal{L}(V, \mathbf{F})$ also gets a special name and special notation:

3.94 Definition ***dual space, V'***

The ***dual space*** of V, denoted V', is the vector space of all linear functionals on V. In other words, $V' = \mathcal{L}(V, \mathbf{F})$.

3.95 $\dim V' = \dim V$

Suppose V is finite-dimensional. Then V' is also finite-dimensional and $\dim V' = \dim V$.

In the following definition, 3.5 implies that each φ_j is well defined.

3.96 Definition ***dual basis***

If $v_1, \dots, v_n$ is a basis of V, then the ***dual basis*** of $v_1, \dots, v_n$ is the list $\varphi_1, \dots, \varphi_n$ of elements of V', where each φ_j is the linear functional on V such that

$$\varphi_j(v_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases}$$

The next result shows that the dual basis is indeed a basis. Thus the terminology "dual basis" is justified.

3.98 Dual basis is a basis of the dual space

Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V'.

In the definition below, note that if T is a linear map from V to W then T' is a linear map from W' to V'.

3.99 Definition ***dual map, T'***

If $T \in \mathcal{L}(V, W)$, then the ***dual map*** of T is the linear map $T' \in \mathcal{L}(W', V')$ defined by $T'(\varphi) = \varphi \circ T$ for $\varphi \in W'$.

If $T \in \mathcal{L}(V, W)$ and $\varphi \in W'$, then $T'(\varphi)$ is defined above to be the composition of the linear maps φ and T. Thus $T'(\varphi)$ is indeed a linear map from V to $\mathbf{F}$; in other words, $T'(\varphi) \in V'$.

The verification that T' is a linear map from W' to V' is easy:

- If $\varphi, \psi \in W'$, then
$$T'(\varphi + \psi) = (\varphi + \psi) \circ T = \varphi \circ T + \psi \circ T = T'(\varphi) + T'(\psi).$$
- If $\lambda \in \mathbf{F}$ and $\varphi \in W'$, then
$$T'(\lambda\varphi) = (\lambda\varphi) \circ T = \lambda(\varphi \circ T) = \lambda T'(\varphi).$$

The first two bullet points in the result below imply that the function that takes T to T' is a linear map from $\mathcal{L}(V, W)$ to $\mathcal{L}(W', V')$.

In the third bullet point below, note the reversal of order from ST on the left to $T'S'$ on the right (here we assume that U is a vector space over $\mathbf{F}$).

3.101 Algebraic properties of dual maps

- $(S + T)' = S' + T'$ for all $S, T \in \mathcal{L}(V, W)$.
- $(\lambda T)' = \lambda T'$ for all $\lambda \in \mathbf{F}$ and all $T \in \mathcal{L}(V, W)$.
- $(ST)' = T'S'$ for all $T \in \mathcal{L}(U, V)$ and all $S \in \mathcal{L}(V, W)$.

The Null Space and Range of the Dual of a Linear Map

Our goal in this subsection is to describe null T' and range T' in terms of range T and null T. To do this, we will need the following definition.

3.102 Definition ***annihilator, U^0***

For $U \subset V$, the ***annihilator*** of U, denoted U^0, is defined by
$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \text{ for all } u \in U\}.$$

For $U \subset V$, the annihilator U^0 is a subset of the dual space V'. Thus U^0 depends on the vector space containing U, so a notation such as U_V^0 would be more precise. However, the containing vector space will always be clear from the context, so we will use the simpler notation U^0.

3.105 The annihilator is a subspace

Suppose $U \subset V$. Then U^0 is a subspace of V'.

The next result shows that $\dim U^0$ is the difference of $\dim V$ and $\dim U$.

3.106 Dimension of the annihilator

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^0 = \dim V.$$

The proof of part (a) of the result below does not use the hypothesis that V and W are finite-dimensional.

3.107 The null space of T'

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

(a) $\operatorname{null} T' = (\operatorname{range} T)^0$;

(b) $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$.

The next result can be useful because sometimes it is easier to verify that T' is injective than to show directly that T is surjective.

3.108 T surjective is equivalent to T' injective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then T is surjective if and only if T' is injective.

3.109 The range of T'

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

(a) $\dim \operatorname{range} T' = \dim \operatorname{range} T$;

(b) $\operatorname{range} T' = (\operatorname{null} T)^0$.

The next result should be compared to 3.108.

3.110 T injective is equivalent to T' surjective

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then T is injective if and only if T' is surjective.

The Matrix of the Dual of a Linear Map

We now define the transpose of a matrix.

3.111 Definition ***transpose,*** A^{t}

The ***transpose*** of a matrix A, denoted A^{t}, is the matrix obtained from A by interchanging the rows and columns. More specifically, if A is an m-by-n matrix, then A^{t} is the n-by-m matrix whose entries are given by the equation

$$(A^{\mathrm{t}})_{k,j} = A_{j,k}.$$

The transpose has nice algebraic properties: $(A + C)^{\mathrm{t}} = A^{\mathrm{t}} + C^{\mathrm{t}}$ and $(\lambda A)^{\mathrm{t}} = \lambda A^{\mathrm{t}}$ for all m-by-n matrices A, C and all $\lambda \in \mathbf{F}$.

The next result shows that the transpose of the product of two matrices is the product of the transposes in the opposite order.

3.113 The transpose of the product of matrices

If A is an m-by-n matrix and C is an n-by-p matrix, then

$$(AC)^{\mathrm{t}} = C^{\mathrm{t}} A^{\mathrm{t}}.$$

The setting for the next result is the assumption that we have a basis $v_1, \dots, v_n$ of V, along with its dual basis $\varphi_1, \dots, \varphi_n$ of V'. We also have a basis $w_1, \dots, w_m$ of W, along with its dual basis $\psi_1, \dots, \psi_m$ of W'. Thus $\mathcal{M}(T)$ is computed with respect to the bases just mentioned of V and W, and $\mathcal{M}(T')$ is computed with respect to the dual bases just mentioned of W' and V'.

3.114 The matrix of T' is the transpose of the matrix of T

Suppose $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(T') = \big(\mathcal{M}(T)\big)^{\mathrm{t}}$.

The Rank of a Matrix

We begin by defining two nonnegative integers that are associated with each matrix.

3.115 Definition ***row rank, column rank***

Suppose A is an m-by-n matrix with entries in $\mathbf{F}$.

- The ***row rank*** of A is the dimension of the span of the rows of A in $\mathbf{F}^{1,n}$.
- The ***column rank*** of A is the dimension of the span of the columns of A in $\mathbf{F}^{m,1}$.

Notice that no bases are in sight in the statement of the next result. Although $\mathcal{M}(T)$ in the next result depends on a choice of bases of V and W, the next result shows that the column rank of $\mathcal{M}(T)$ is the same for all such choices (because range T does not depend on a choice of basis).

3.117 Dimension of range T equals column rank of $\mathcal{M}(T)$

Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then $\dim \operatorname{range} T$ equals the column rank of $\mathcal{M}(T)$.

3.118 Row rank equals column rank

Suppose $A \in \mathbf{F}^{m,n}$. Then the row rank of A equals the column rank of A.

The last result allows us to dispense with the terms “row rank” and “column rank” and just use the simpler term “rank”.

3.119 Definition ***rank***

The ***rank*** of a matrix $A \in \mathbf{F}^{m,n}$ is the column rank of A.

CHAPTER

4

Statue of Persian mathematician and poet Omar Khayyám (1048–1131), whose algebra book written in 1070 contained the first serious study of cubic polynomials.

Polynomials

This short chapter contains material on polynomials that we will need to understand operators. Many of the results in this chapter will already be familiar to you from other courses; they are included here for completeness.

Because this chapter is not about linear algebra, your instructor may go through it rapidly. You may not be asked to scrutinize all the proofs. Make sure, however, that you at least read and understand the statements of all the results in this chapter—they will be used in later chapters.

The standing assumption we need for this chapter is as follows:

4.1 Notation $\mathbf{F}$

$\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- Division Algorithm for Polynomials
- factorization of polynomials over $\mathbf{C}$
- factorization of polynomials over $\mathbf{R}$

Linear Algebra Abridged is generated from *Linear Algebra Done Right* (by Sheldon Axler, third edition) by excluding all proofs, examples, and exercises, along with most comments. The full version of *Linear Algebra Done Right* is available at springer.com and amazon.com in both printed and electronic forms.

Complex Conjugate and Absolute Value

Before discussing polynomials with complex or real coefficients, we need to learn a bit more about the complex numbers.

4.2 Definition Re z, Im z

Suppose $z = a + bi$, where a and b are real numbers.

- The ***real part*** of z, denoted $\operatorname{Re} z$, is defined by $\operatorname{Re} z = a$.
- The ***imaginary part*** of z, denoted $\operatorname{Im} z$, is defined by $\operatorname{Im} z = b$.

Thus for every complex number z, we have

$$z = \operatorname{Re} z + (\operatorname{Im} z)i.$$

4.3 Definition ***complex conjugate, $\bar{z}$, absolute value, $|z|$***

Suppose $z \in \mathbf{C}$.

- The ***complex conjugate*** of $z \in \mathbf{C}$, denoted $\bar{z}$, is defined by
$$\bar{z} = \operatorname{Re} z - (\operatorname{Im} z)i.$$
- The ***absolute value*** of a complex number z, denoted $|z|$, is defined by
$$|z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

Note that $|z|$ is a nonnegative number for every $z \in \mathbf{C}$.

The real and imaginary parts, complex conjugate, and absolute value have the following properties:

4.5 Properties of complex numbers

Suppose $w, z \in \mathbf{C}$. Then

sum of z and $\bar{z}$

$z + \bar{z} = 2 \operatorname{Re} z;$

difference of z and $\bar{z}$

$z - \bar{z} = 2(\operatorname{Im} z)i;$

product of z and $\bar{z}$

$z\bar{z} = |z|^2;$

additivity and multiplicativity of complex conjugate

$\overline{w + z} = \bar{w} + \bar{z}$ and $\overline{wz} = \bar{w}\bar{z};$

conjugate of conjugate

$\bar{\bar{z}} = z;$

real and imaginary parts are bounded by $|z|$

$|\operatorname{Re} z| \leq |z|$ and $|\operatorname{Im} z| \leq |z|$

absolute value of the complex conjugate

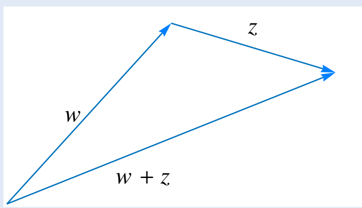
$|\bar{z}| = |z|;$

multiplicativity of absolute value

$|wz| = |w|\,|z|;$

Triangle Inequality

$|w + z| \leq |w| + |z|.$



Uniqueness of Coefficients for Polynomials

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

4.6
$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

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

4.7 If a polynomial is the zero function, then all coefficients are 0

Suppose $a_0, \ldots, a_m \in \mathbf{F}$. If

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

for every $z \in \mathbf{F}$, then $a_0 = \cdots = a_m = 0$.

The result above implies that the coefficients of a polynomial are uniquely determined (because if a polynomial had two different sets of coefficients, then subtracting the two representations of the polynomial would give a contradiction to the result above).

Recall that if a polynomial p can be written in the form 4.6 with $a_m \neq 0$, then we say that p has degree m and we write $\deg p = m$.

The degree of the 0 polynomial is defined to be $-\infty$. When necessary, use the obvious arithmetic with $-\infty$. For example, $-\infty < m$ and $-\infty + m = -\infty$ for every integer m.

The Division Algorithm for Polynomials

If p and s are nonnegative integers, with $s \neq 0$, then there exist nonnegative integers q and r such that

$$p = sq + r$$

and $r < s$. Think of dividing p by s, getting quotient q with remainder r. Our next task is to prove an analogous result for polynomials.

The result below is often called the Division Algorithm for Polynomials, although as stated here it is not really an algorithm, just a useful result.

Recall that $\mathcal{P}(\mathbf{F})$ denotes the vector space of all polynomials with coefficients in $\mathbf{F}$ and that $\mathcal{P}_m(\mathbf{F})$ is the subspace of $\mathcal{P}(\mathbf{F})$ consisting of the polynomials with coefficients in $\mathbf{F}$ and degree at most m.

The next result can be proved without linear algebra, but the proof given here using linear algebra is appropriate for a linear algebra textbook.

4.8 Division Algorithm for Polynomials

Suppose that $p, s \in \mathcal{P}(\mathbf{F})$, with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathcal{P}(\mathbf{F})$ such that

$$p = sq + r$$

and $\deg r < \deg s$.

Zeros of Polynomials

The solutions to the equation $p(z) = 0$ play a crucial role in the study of a polynomial $p \in \mathcal{P}(\mathbf{F})$. Thus these solutions have a special name.

4.9 Definition ***zero of a polynomial***

A number $\lambda \in \mathbf{F}$ is called a ***zero*** (or ***root***) of a polynomial $p \in \mathcal{P}(\mathbf{F})$ if

$$p(\lambda) = 0.$$

4.10 Definition ***factor***

A polynomial $s \in \mathcal{P}(\mathbf{F})$ is called a ***factor*** of $p \in \mathcal{P}(\mathbf{F})$ if there exists a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that $p = sq$.

We begin by showing that λ is a zero of a polynomial $p \in \mathcal{P}(\mathbf{F})$ if and only if $z - \lambda$ is a factor of p.

4.11 Each zero of a polynomial corresponds to a degree-1 factor

Suppose $p \in \mathcal{P}(\mathbf{F})$ and $\lambda \in \mathbf{F}$. Then $p(\lambda) = 0$ if and only if there is a polynomial $q \in \mathcal{P}(\mathbf{F})$ such that

$$p(z) = (z - \lambda)q(z)$$

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

Now we can prove that polynomials do not have too many zeros.

4.12 A polynomial has at most as many zeros as its degree

Suppose $p \in \mathcal{P}(\mathbf{F})$ is a polynomial with degree $m \geq 0$. Then p has at most m distinct zeros in $\mathbf{F}$.

Factorization of Polynomials over C

So far we have been handling polynomials with complex coefficients and polynomials with real coefficients simultaneously through our convention that $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$. Now we will see some differences between these two cases. First we treat polynomials with complex coefficients. Then we will use our results about polynomials with complex coefficients to prove corresponding results for polynomials with real coefficients.

The next result, although called the Fundamental Theorem of Algebra, uses analysis in its proof. The short proof presented here uses tools from complex analysis. If you have not had a course in complex analysis, this proof will almost certainly be meaningless to you. In that case, just accept the Fundamental Theorem of Algebra as something that we need to use but whose proof requires more advanced tools that you may learn in later courses.

4.13 Fundamental Theorem of Algebra

Every nonconstant polynomial with complex coefficients has a zero.

Although the proof given above is probably the shortest proof of the Fundamental Theorem of Algebra, a web search can lead you to several other proofs that use different techniques. All proofs of the Fundamental Theorem of Algebra need to use some analysis, because the result is not true if $\mathbf{C}$ is replaced, for example, with the set of numbers of the form $c + di$ where c, d are rational numbers.

Remarkably, mathematicians have proved that no formula exists for the zeros of polynomials of degree 5 or higher. But computers and calculators can use clever numerical methods to find good approximations to the zeros of any polynomial, even when exact zeros cannot be found.

For example, no one will ever be able to give an exact formula for a zero of the polynomial p defined by

$$p(x) = x^5 - 5x^4 - 6x^3 + 17x^2 + 4x - 7.$$

However, a computer or symbolic calculator can find approximate zeros of this polynomial.

The Fundamental Theorem of Algebra leads to the following factorization result for polynomials with complex coefficients. Note that in this factorization, the numbers $\lambda_1, \ldots, \lambda_m$ are precisely the zeros of p, for these are the only values of z for which the right side of the equation in the next result equals 0.

4.14 Factorization of a polynomial over $\mathbf{C}$

If $p \in \mathcal{P}(\mathbf{C})$ is a nonconstant polynomial, then p has a unique factorization (except for the order of the factors) of the form

$$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m),$$

where $c, \lambda_1, \ldots, \lambda_m \in \mathbf{C}$.

Factorization of Polynomials over R

A polynomial with real coefficients may have no real zeros. For example, the polynomial $1 + x^2$ has no real zeros.

To obtain a factorization theorem over **R**, we will use our factorization theorem over **C**. We begin with the following result.

4.15 Polynomials with real coefficients have zeros in pairs

Suppose $p \in \mathcal{P}(\mathbf{C})$ is a polynomial with real coefficients. If $\lambda \in \mathbf{C}$ is a zero of p, then so is $\bar{\lambda}$.

We want a factorization theorem for polynomials with real coefficients. First we need to characterize the polynomials of degree 2 with real coefficients that can be written as the product of two polynomials of degree 1 with real coefficients.

4.16 Factorization of a quadratic polynomial

Suppose $b, c \in \mathbf{R}$. Then there is a polynomial factorization of the form

$$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

with $\lambda_1, \lambda_2 \in \mathbf{R}$ if and only if $b^2 \geq 4c$.

The next result gives a factorization of a polynomial over **R**. The idea of the proof is to use the factorization 4.14 of p as a polynomial with complex coefficients. Complex but nonreal zeros of p come in pairs; see 4.15. Thus if the factorization of p as an element of $\mathcal{P}(\mathbf{C})$ includes terms of the form $(x - \lambda)$ with λ a nonreal complex number, then $(x - \bar{\lambda})$ is also a term in the factorization. Multiplying together these two terms, we get

$$\big(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\big),$$

which is a quadratic term of the required form.

The idea sketched in the paragraph above almost provides a proof of the existence of our desired factorization. However, we need to be careful about one point. Suppose λ is a nonreal complex number and $(x - \lambda)$ is a term in the factorization of p as an element of $\mathcal{P}(\mathbf{C})$. We are guaranteed by 4.15 that $(x - \bar{\lambda})$ also appears as a term in the factorization, but 4.15 does not state that these two factors appear the same number of times, as needed to make the idea above work. However, the proof works around this point.

In the next result, either m or M may equal 0. The numbers $\lambda_1, \ldots, \lambda_m$ are precisely the real zeros of p, for these are the only real values of x for which the right side of the equation in the next result equals 0.

4.17 Factorization of a polynomial over $\mathbf{R}$

Suppose $p \in \mathcal{P}(\mathbf{R})$ is a nonconstant polynomial. Then p has a unique factorization (except for the order of the factors) of the form

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

where $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbf{R}$, with ${b_j}^2 < 4c_j$ for each j.

CHAPTER

5

Statue of Italian mathematician Leonardo of Pisa (1170–1250, approximate dates), also known as Fibonacci.

Eigenvalues, Eigenvectors, and Invariant Subspaces

Linear maps from one vector space to another vector space were the objects of study in Chapter 3. Now we begin our investigation of linear maps from a finite-dimensional vector space to itself. Their study constitutes the most important part of linear algebra.

Our standing assumptions are as follows:

5.1 **Notation** $\mathbf{F}$, V

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

LEARNING OBJECTIVES FOR THIS CHAPTER

- invariant subspaces
- eigenvalues, eigenvectors, and eigenspaces
- each operator on a finite-dimensional complex vector space has an eigenvalue and an upper-triangular matrix with respect to some basis

5.A Invariant Subspaces

In this chapter we develop the tools that will help us understand the structure of operators. Recall that an operator is a linear map from a vector space to itself. Recall also that we denote the set of operators on V by $\mathcal{L}(V)$; in other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

Let's see how we might better understand what an operator looks like. Suppose $T \in \mathcal{L}(V)$. If we have a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a proper subspace of V, then to understand the behavior of T, we need only understand the behavior of each $T|_{U_j}$; here $T|_{U_j}$ denotes the restriction of T to the smaller domain U_j. Dealing with $T|_{U_j}$ should be easier than dealing with T because U_j is a smaller vector space than V.

However, if we intend to apply tools useful in the study of operators (such as taking powers), then we have a problem: $T|_{U_j}$ may not map U_j into itself; in other words, $T|_{U_j}$ may not be an operator on U_j. Thus we are led to consider only decompositions of V of the form above where T maps each U_j into itself.

The notion of a subspace that gets mapped into itself is sufficiently important to deserve a name.

5.2 Definition ***invariant subspace***

Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called ***invariant*** under T if $u \in U$ implies $Tu \in U$.

In other words, U is invariant under T if $T|_U$ is an operator on U.

Must an operator $T \in \mathcal{L}(V)$ have any invariant subspaces other than $\{0\}$ and V? Later we will see that this question has an affirmative answer if V is finite-dimensional and $\dim V > 1$ (for $\mathbf{F} = \mathbf{C}$) or $\dim V > 2$ (for $\mathbf{F} = \mathbf{R}$); see 5.21 and 9.8.

Although null T and range T are invariant under T, they do not necessarily provide easy answers to the question about the existence of invariant subspaces other than $\{0\}$ and V, because null T may equal $\{0\}$ and range T may equal V (this happens when T is invertible).

Eigenvalues and Eigenvectors

We will return later to a deeper study of invariant subspaces. Now we turn to an investigation of the simplest possible nontrivial invariant subspaces—invariant subspaces with dimension 1.

Take any $v \in V$ with $v \neq 0$ and let U equal the set of all scalar multiples of v:

$$U = \{\lambda v : \lambda \in \mathbf{F}\} = \operatorname{span}(v).$$

Then U is a 1-dimensional subspace of V (and every 1-dimensional subspace of V is of this form for an appropriate choice of v). If U is invariant under an operator $T \in \mathcal{L}(V)$, then $Tv \in U$, and hence there is a scalar $\lambda \in \mathbf{F}$ such that

$$Tv = \lambda v.$$

Conversely, if $Tv = \lambda v$ for some $\lambda \in \mathbf{F}$, then $\operatorname{span}(v)$ is a 1-dimensional subspace of V invariant under T.

The equation

$$Tv = \lambda v,$$

which we have just seen is intimately connected with 1-dimensional invariant subspaces, is important enough that the vectors v and scalars λ satisfying it are given special names.

5.5 Definition ***eigenvalue***

Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbf{F}$ is called an ***eigenvalue*** of T if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

The comments above show that T has a 1-dimensional invariant subspace if and only if T has an eigenvalue.

In the definition above, we require that $v \neq 0$ because every scalar $\lambda \in \mathbf{F}$ satisfies $T0 = \lambda 0$.

5.6 Equivalent conditions to be an eigenvalue

Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $\lambda \in F$. Then the following are equivalent:

(a) λ is an eigenvalue of T;

(b) $T - \lambda I$ is not injective;

(c) $T - \lambda I$ is not surjective;

(d) $T - \lambda I$ is not invertible.

Recall that $I \in \mathcal{L}(V)$ is the identity operator defined by $Iv = v$ for all $v \in V$.

5.7 Definition ***eigenvector***

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$ is an eigenvalue of T. A vector $v \in V$ is called an ***eigenvector*** of T corresponding to λ if $v \neq 0$ and $Tv = \lambda v$.

Because $Tv = \lambda v$ if and only if $(T - \lambda I)v = 0$, a vector $v \in V$ with $v \neq 0$ is an eigenvector of T corresponding to λ if and only if $v \in \operatorname{null}(T - \lambda I)$.

Now we show that eigenvectors corresponding to distinct eigenvalues are linearly independent.

5.10 Linearly independent eigenvectors

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.

The corollary below states that an operator cannot have more distinct eigenvalues than the dimension of the vector space on which it acts.

5.13 Number of eigenvalues

Suppose V is finite-dimensional. Then each operator on V has at most $\dim V$ distinct eigenvalues.

Restriction and Quotient Operators

If $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T, then U determines two other operators $T|_U \in \mathcal{L}(U)$ and $T/U \in \mathcal{L}(V/U)$ in a natural way, as defined below.

5.14 Definition $T|_U$ ***and*** T/U

Suppose $T \in \mathcal{L}(V)$ and U is a subspace of V invariant under T.

- The ***restriction operator*** $T|_U \in \mathcal{L}(U)$ is defined by
$$T|_U(u) = Tu$$
for $u \in U$.
- The ***quotient operator*** $T/U \in \mathcal{L}(V/U)$ is defined by
$$(T/U)(v + U) = Tv + U$$
for $v \in V$.

For both the operators defined above, it is worthwhile to pay attention to their domains and to spend a moment thinking about why they are well defined as operators on their domains. First consider the restriction operator $T|_U \in \mathcal{L}(U)$, which is T with its domain restricted to U, thought of as mapping into U instead of into V. The condition that U is invariant under T is what allows us to think of $T|_U$ as an operator on U, meaning a linear map into the same space as the domain, rather than as simply a linear map from one vector space to another vector space.

To show that the definition above of the quotient operator makes sense, we need to verify that if $v + U = w + U$, then $Tv + U = Tw + U$. Hence suppose $v + U = w + U$. Thus $v - w \in U$ (see 3.85). Because U is invariant under T, we also have $T(v - w) \in U$, which implies that $Tv - Tw \in U$, which implies that $Tv + U = Tw + U$, as desired.

Suppose T is an operator on a finite-dimensional vector space V and U is a subspace of V invariant under T, with $U \neq \{0\}$ and $U \neq V$. In some sense, we can learn about T by studying the operators $T|_U$ and T/U, each of which is an operator on a vector space with smaller dimension than V.

5.B Eigenvectors and Upper-Triangular Matrices

Polynomials Applied to Operators

The main reason that a richer theory exists for operators (which map a vector space into itself) than for more general linear maps is that operators can be raised to powers.

If $T \in \mathcal{L}(V)$, then TT makes sense and is also in $\mathcal{L}(V)$. We usually write T^2 instead of TT. More generally, we have the following definition.

5.16 Definition T^m

Suppose $T \in \mathcal{L}(V)$ and m is a positive integer.

- T^m is defined by
$$T^m = \underbrace{T \cdots T}_{m \text{ times}}.$$
- T^0 is defined to be the identity operator I on V.
- If T is invertible with inverse T^{-1}, then T^{-m} is defined by
$$T^{-m} = (T^{-1})^m.$$

You should verify that if T is an operator, then

$$T^m T^n = T^{m+n} \quad \text{and} \quad (T^m)^n = T^{mn},$$

where m and n are allowed to be arbitrary integers if T is invertible and nonnegative integers if T is not invertible.

5.17 Definition $p(T)$

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$ is a polynomial given by

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

for $z \in \mathbf{F}$. Then $p(T)$ is the operator defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

This is a new use of the symbol p because we are applying it to operators, not just elements of $\mathbf{F}$.

If we fix an operator $T \in \mathcal{L}(V)$, then the function from $\mathcal{P}(\mathbf{F})$ to $\mathcal{L}(V)$ given by $p \mapsto p(T)$ is linear, as you should verify.

5.19 Definition ***product of polynomials***

If $p, q \in \mathcal{P}(\mathbf{F})$, then $pq \in \mathcal{P}(\mathbf{F})$ is the polynomial defined by

$$(pq)(z) = p(z)q(z)$$

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

Any two polynomials of an operator commute, as shown below.

5.20 Multiplicative properties

Suppose $p, q \in \mathcal{P}(\mathbf{F})$ and $T \in \mathcal{L}(V)$. Then

(a) $(pq)(T) = p(T)q(T)$;

(b) $p(T)q(T) = q(T)p(T)$.

Part (a) holds because when expanding a product of polynomials using the distributive property, it does not matter whether the symbol is z or T.

Existence of Eigenvalues

Now we come to one of the central results about operators on complex vector spaces.

5.21 Operators on complex vector spaces have an eigenvalue

Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.

Upper-Triangular Matrices

In Chapter 3 we discussed the matrix of a linear map from one vector space to another vector space. That matrix depended on a choice of a basis of each of the two vector spaces. Now that we are studying operators, which map a vector space to itself, the emphasis is on using only one basis.

5.22 Definition ***matrix of an operator, $\mathcal{M}(T)$***

Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of V. The ***matrix of*** T with respect to this basis is the n-by-n matrix

$$\mathcal{M}(T) = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}$$

whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}v_1 + \dots + A_{n,k}v_n.$$

If the basis is not clear from the context, then the notation $\mathcal{M}\big(T, (v_1, \dots, v_n)\big)$ is used.

Note that the matrices of operators are square arrays, rather than the more general rectangular arrays that we considered earlier for linear maps.

If T is an operator on $\mathbf{F}^n$ and no basis is specified, assume that the basis in question is the standard one (where the j^{th} basis vector is 1 in the j^{th} slot and 0 in all the other slots). You can then think of the j^{th} column of $\mathcal{M}(T)$ as T applied to the j^{th} basis vector.

A central goal of linear algebra is to show that given an operator $T \in \mathcal{L}(V)$, there exists a basis of V with respect to which T has a reasonably simple matrix. To make this vague formulation a bit more precise, we might try to choose a basis of V such that $\mathcal{M}(T)$ has many 0's.

If V is a finite-dimensional complex vector space, then we already know enough to show that there is a basis of V with respect to which the matrix of T has 0's everywhere in the first column, except possibly the first entry. In other words, there is a basis of V with respect to which the matrix of T looks like

$$\begin{pmatrix} \lambda & & \\ 0 & * & \\ \vdots & & \\ 0 & & \end{pmatrix};$$

here the $*$ denotes the entries in all the columns other than the first column. To prove this, let λ be an eigenvalue of T (one exists by 5.21) and let v be a corresponding eigenvector. Extend v to a basis of V. Then the matrix of T with respect to this basis has the form above.

Soon we will see that we can choose a basis of V with respect to which the matrix of T has even more 0's.

5.24 Definition ***diagonal of a matrix***

The ***diagonal*** of a square matrix consists of the entries along the line from the upper left corner to the bottom right corner.

5.25 Definition ***upper-triangular matrix***

A matrix is called ***upper triangular*** if all the entries below the diagonal equal 0.

Typically we represent an upper-triangular matrix in the form

$$\begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix};$$

the 0 in the matrix above indicates that all entries below the diagonal in this n-by-n matrix equal 0. Upper-triangular matrices can be considered reasonably simple—for n large, almost half its entries in an n-by-n upper-triangular matrix are 0.

The following proposition demonstrates a useful connection between upper-triangular matrices and invariant subspaces.

5.26 Conditions for upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following are equivalent:

(a) the matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular;

(b) $T v_j \in \operatorname{span}(v_1, \ldots, v_j)$ for each $j = 1, \ldots, n$;

(c) $\operatorname{span}(v_1, \ldots, v_j)$ is invariant under T for each $j = 1, \ldots, n$.

Now we can prove that for each operator on a finite-dimensional complex vector space, there is a basis of the vector space with respect to which the matrix of the operator has only 0's below the diagonal. In Chapter 8 we will improve even this result.

Sometimes more insight comes from seeing more than one proof of a theorem. Thus two proofs are presented of the next result. Use whichever appeals more to you.

5.27 Over $\mathbf{C}$, every operator has an upper-triangular matrix

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.

How does one determine from looking at the matrix of an operator whether the operator is invertible? If we are fortunate enough to have a basis with respect to which the matrix of the operator is upper triangular, then this problem becomes easy, as the following proposition shows.

5.30 Determination of invertibility from upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.

Unfortunately no method exists for exactly computing the eigenvalues of an operator from its matrix. However, if we are fortunate enough to find a basis with respect to which the matrix of the operator is upper triangular, then the problem of computing the eigenvalues becomes trivial, as the following proposition shows.

5.32 Determination of eigenvalues from upper-triangular matrix

Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

Once the eigenvalues of an operator on $\mathbf{F}^n$ are known, the eigenvectors can be found easily using Gaussian elimination.

5.C *Eigenspaces and Diagonal Matrices*

5.34 Definition ***diagonal matrix***

A ***diagonal matrix*** is a square matrix that is 0 everywhere except possibly along the diagonal.

Obviously every diagonal matrix is upper triangular. In general, a diagonal matrix has many more 0's than an upper-triangular matrix.

If an operator has a diagonal matrix with respect to some basis, then the entries along the diagonal are precisely the eigenvalues of the operator; this follows from 5.32 (or find an easier proof for diagonal matrices).

5.36 Definition ***eigenspace,*** $E(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The ***eigenspace*** of T corresponding to λ, denoted $E(\lambda, T)$, is defined by

$$E(\lambda, T) = \operatorname{null}(T - \lambda I).$$

In other words, $E(\lambda, T)$ is the set of all eigenvectors of T corresponding to λ, along with the 0 vector.

For $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, the eigenspace $E(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V). The definitions imply that λ is an eigenvalue of T if and only if $E(\lambda, T) \neq \{0\}$.

If λ is an eigenvalue of an operator $T \in \mathcal{L}(V)$, then T restricted to $E(\lambda, T)$ is just the operator of multiplication by λ.

5.38 Sum of eigenspaces is a direct sum

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum. Furthermore,

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

5.39 Definition ***diagonalizable***

An operator $T \in \mathcal{L}(V)$ is called ***diagonalizable*** if the operator has a diagonal matrix with respect to some basis of V.

5.41 Conditions equivalent to diagonalizability

Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T. Then the following are equivalent:

(a) T is diagonalizable;

(b) V has a basis consisting of eigenvectors of T;

(c) there exist 1-dimensional subspaces $U_1, \dots, U_n$ of V, each invariant under T, such that

$$V = U_1 \oplus \cdots \oplus U_n;$$

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

(e) $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Unfortunately not every operator is diagonalizable. This sad state of affairs can arise even on complex vector spaces.

The next result shows that if an operator has as many distinct eigenvalues as the dimension of its domain, then the operator is diagonalizable.

5.44 Enough eigenvalues implies diagonalizability

If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.

The converse of 5.44 is not true. For example, the operator T defined on the three-dimensional space $\mathbf{F}^3$ by

$$T(z_1, z_2, z_3) = (4z_1, 4z_2, 5z_3)$$

has only two eigenvalues (4 and 5), but this operator has a diagonal matrix with respect to the standard basis.

In later chapters we will find additional conditions that imply that certain operators are diagonalizable.

CHAPTER

6

Woman teaching geometry, from a fourteenth-century edition of Euclid's geometry book.

Inner Product Spaces

In making the definition of a vector space, we generalized the linear structure (addition and scalar multiplication) of $\mathbf{R}^2$ and $\mathbf{R}^3$. We ignored other important features, such as the notions of length and angle. These ideas are embedded in the concept we now investigate, inner products.

Our standing assumptions are as follows:

6.1 Notation $\mathbf{F}, V$

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

LEARNING OBJECTIVES FOR THIS CHAPTER

- Cauchy–Schwarz Inequality
- Gram–Schmidt Procedure
- linear functionals on inner product spaces
- calculating minimum distance to a subspace

Linear Algebra Abridged is generated from *Linear Algebra Done Right* (by Sheldon Axler, third edition) by excluding all proofs, examples, and exercises, along with most comments. The full version of *Linear Algebra Done Right* is available at springer.com and amazon.com in both printed and electronic forms.

6.A Inner Products and Norms

Inner Products

To motivate the concept of inner product, think of vectors in $\mathbf{R}^2$ and $\mathbf{R}^3$ as arrows with initial point at the origin. The length of a vector x in $\mathbf{R}^2$ or $\mathbf{R}^3$ is called the ***norm*** of x, denoted $\|x\|$. Thus for $x = (x_1, x_2) \in \mathbf{R}^2$, we have $\|x\| = \sqrt{x_1{}^2 + x_2{}^2}$.

Similarly, if $x = (x_1, x_2, x_3) \in \mathbf{R}^3$, then $\|x\| = \sqrt{x_1{}^2 + x_2{}^2 + x_3{}^2}$.

Even though we cannot draw pictures in higher dimensions, the generalization to $\mathbf{R}^n$ is obvious: we define the norm of $x = (x_1, \dots, x_n) \in \mathbf{R}^n$ by

$$\|x\| = \sqrt{x_1{}^2 + \cdots + x_n{}^2}.$$

The norm is not linear on $\mathbf{R}^n$. To inject linearity into the discussion, we introduce the dot product.

6.2 Definition ***dot product***

For $x, y \in \mathbf{R}^n$, the ***dot product*** of x and y, denoted $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$.

Note that the dot product of two vectors in $\mathbf{R}^n$ is a number, not a vector. Obviously $x \cdot x = \|x\|^2$ for all $x \in \mathbf{R}^n$. The dot product on $\mathbf{R}^n$ has the following properties:

- $x \cdot x \geq 0$ for all $x \in \mathbf{R}^n$;
- $x \cdot x = 0$ if and only if $x = 0$;
- for $y \in \mathbf{R}^n$ fixed, the map from $\mathbf{R}^n$ to $\mathbf{R}$ that sends $x \in \mathbf{R}^n$ to $x \cdot y$ is linear;
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbf{R}^n$.

An inner product is a generalization of the dot product. At this point you may be tempted to guess that an inner product is defined by abstracting the properties of the dot product discussed in the last paragraph. For real vector spaces, that guess is correct. However, so that we can make a definition that will be useful for both real and complex vector spaces, we need to examine the complex case before making the definition.

Recall that if $\lambda = a + bi$, where $a, b \in \mathbf{R}$, then

- the absolute value of λ, denoted $|\lambda|$, is defined by $|\lambda| = \sqrt{a^2 + b^2}$;
- the complex conjugate of λ, denoted $\bar{\lambda}$, is defined by $\bar{\lambda} = a - bi$;
- $|\lambda|^2 = \lambda\bar{\lambda}$.

See Chapter 4 for the definitions and the basic properties of the absolute value and complex conjugate.

For $z = (z_1, \dots, z_n) \in \mathbf{C}^n$, we define the norm of z by

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

The absolute values are needed because we want $\|z\|$ to be a nonnegative number. Note that

$$\|z\|^2 = z_1\overline{z_1} + \cdots + z_n\overline{z_n}.$$

We want to think of $\|z\|^2$ as the inner product of z with itself, as we did in $\mathbf{R}^n$. The equation above thus suggests that the inner product of $w = (w_1, \dots, w_n) \in \mathbf{C}^n$ with z should equal

$$w_1\overline{z_1} + \cdots + w_n\overline{z_n}.$$

If the roles of the w and z were interchanged, the expression above would be replaced with its complex conjugate. In other words, we should expect that the inner product of w with z equals the complex conjugate of the inner product of z with w. With that motivation, we are now ready to define an inner product on V, which may be a real or a complex vector space.

Two comments about the notation used in the next definition:

- If λ is a complex number, then the notation $\lambda \geq 0$ means that λ is real and nonnegative.
- We use the common notation $\langle u, v \rangle$, with angle brackets denoting an inner product. Some people use parentheses instead, but then (u, v) becomes ambiguous because it could denote either an ordered pair or an inner product.

6.3 Definition ***inner product***

An ***inner product*** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v\rangle \in \mathbf{F}$ and has the following properties:

positivity
$\langle v, v\rangle \geq 0$ for all $v \in V$;

definiteness
$\langle v, v\rangle = 0$ if and only if $v = 0$;

additivity in first slot
$\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$;

homogeneity in first slot
$\langle \lambda u, v\rangle = \lambda\langle u, v\rangle$ for all $\lambda \in \mathbf{F}$ and all $u, v \in V$;

conjugate symmetry
$\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$.

Every real number equals its complex conjugate. Thus if we are dealing with a real vector space, then in the last condition above we can dispense with the complex conjugate and simply state that $\langle u, v\rangle = \langle v, u\rangle$ for all $v, w \in V$.

6.5 Definition ***inner product space***

An ***inner product space*** is a vector space V along with an inner product on V.

The most important example of an inner product space is $\mathbf{F}^n$ with the Euclidean inner product given by part (a) of the last example. When $\mathbf{F}^n$ is referred to as an inner product space, you should assume that the inner product is the Euclidean inner product unless explicitly told otherwise.

So that we do not have to keep repeating the hypothesis that V is an inner product space, for the rest of this chapter we make the following assumption:

6.6 Notation V

For the rest of this chapter, V denotes an inner product space over $\mathbf{F}$.

Note the slight abuse of language here. An inner product space is a vector space along with an inner product on that vector space. When we say that a vector space V is an inner product space, we are also thinking that an inner product on V is lurking nearby or is obvious from the context (or is the Euclidean inner product if the vector space is $\mathbf{F}^n$).

6.7 Basic properties of an inner product

(a) For each fixed $u \in V$, the function that takes v to $\langle v, u \rangle$ is a linear map from V to $\mathbf{F}$.

(b) $\langle 0, u \rangle = 0$ for every $u \in V$.

(c) $\langle u, 0 \rangle = 0$ for every $u \in V$.

(d) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u, v, w \in V$.

(e) $\langle u, \lambda v \rangle = \bar{\lambda} \langle u, v \rangle$ for all $\lambda \in \mathbf{F}$ and $u, v \in V$.

Norms

Our motivation for defining inner products came initially from the norms of vectors on $\mathbf{R}^2$ and $\mathbf{R}^3$. Now we see that each inner product determines a norm.

6.8 Definition ***norm,*** $\|v\|$

For $v \in V$, the ***norm*** of v, denoted $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

6.10 Basic properties of the norm

Suppose $v \in V$.

(a) $\|v\| = 0$ if and only if $v = 0$.

(b) $\|\lambda v\| = |\lambda| \, \|v\|$ for all $\lambda \in \mathbf{F}$.

The proof above of part (b) illustrates a general principle: working with norms squared is usually easier than working directly with norms.

Now we come to a crucial definition.

6.11 Definition ***orthogonal***

Two vectors $u, v \in V$ are called ***orthogonal*** if $\langle u, v \rangle = 0$.

In the definition above, the order of the vectors does not matter, because $\langle u, v \rangle = 0$ if and only if $\langle v, u \rangle = 0$. Instead of saying that u and v are orthogonal, sometimes we say that u is orthogonal to v.

We begin our study of orthogonality with an easy result.

6.12 Orthogonality and 0

(a) 0 is orthogonal to every vector in V.

(b) 0 is the only vector in V that is orthogonal to itself.

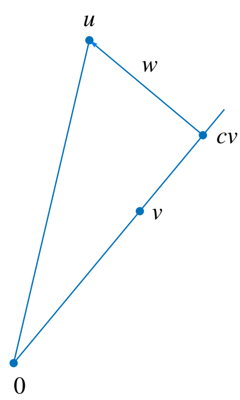
For the special case $V = \mathbf{R}^2$, the next theorem is over 2,500 years old. Of course, the proof below is not the original proof.

6.13 Pythagorean Theorem

Suppose u and v are orthogonal vectors in V. Then

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Suppose $u, v \in V$, with $v \neq 0$. We would like to write u as a scalar multiple of v plus a vector w orthogonal to v, as suggested in the next picture.



An orthogonal decomposition.

To discover how to write u as a scalar multiple of v plus a vector orthogonal to v, let $c \in \mathbf{F}$ denote a scalar. Then

$$u = cv + (u - cv).$$

Thus we need to choose c so that v is orthogonal to $(u - cv)$. In other words, we want

$$0 = \langle u - cv, v \rangle = \langle u, v \rangle - c\|v\|^2.$$

The equation above shows that we should choose c to be $\langle u, v \rangle / \|v\|^2$. Making this choice of c, we can write

$$u = \frac{\langle u, v \rangle}{\|v\|^2} v + \left(u - \frac{\langle u, v \rangle}{\|v\|^2} v \right).$$

As you should verify, the equation above writes u as a scalar multiple of v plus a vector orthogonal to v. In other words, we have proved the following result.

6.14 An orthogonal decomposition

Suppose $u, v \in V$, with $v \neq 0$. Set $c = \dfrac{\langle u, v \rangle}{\|v\|^2}$ and $w = u - \dfrac{\langle u, v \rangle}{\|v\|^2} v$. Then

$$\langle w, v \rangle = 0 \quad \text{and} \quad u = cv + w.$$

The orthogonal decomposition 6.14 will be used in the proof of the Cauchy–Schwarz Inequality, which is our next result and is one of the most important inequalities in mathematics.

6.15 Cauchy–Schwarz Inequality

Suppose $u, v \in V$. Then

$$|\langle u, v \rangle| \leq \|u\| \, \|v\|.$$

This inequality is an equality if and only if one of u, v is a scalar multiple of the other.

The next result, called the Triangle Inequality, has the geometric interpretation that the length of each side of a triangle is less than the sum of the lengths of the other two sides.

Note that the Triangle Inequality implies that the shortest path between two points is a line segment.

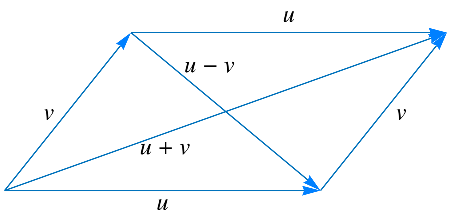
6.18 Triangle Inequality

Suppose $u, v \in V$. Then

$$\|u + v\| \leq \|u\| + \|v\|.$$

This inequality is an equality if and only if one of u, v is a nonnegative multiple of the other.

The next result is called the parallelogram equality because of its geometric interpretation: in every parallelogram, the sum of the squares of the lengths of the diagonals equals the sum of the squares of the lengths of the four sides.



The parallelogram equality.

6.22 Parallelogram Equality

Suppose $u, v \in V$. Then

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

6.B *Orthonormal Bases*

6.23 Definition ***orthonormal***

- A list of vectors is called ***orthonormal*** if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.
- In other words, a list $e_1, \ldots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Orthonormal lists are particularly easy to work with, as illustrated by the next result.

6.25 The norm of an orthonormal linear combination

If $e_1, \dots, e_m$ is an orthonormal list of vectors in V, then

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

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

The result above has the following important corollary.

6.26 An orthonormal list is linearly independent

Every orthonormal list of vectors is linearly independent.

6.27 Definition ***orthonormal basis***

An ***orthonormal basis*** of V is an orthonormal list of vectors in V that is also a basis of V.

For example, the standard basis is an orthonormal basis of $\mathbf{F}^n$.

6.28 An orthonormal list of the right length is an orthonormal basis

Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.

In general, given a basis $e_1, \dots, e_n$ of V and a vector $v \in V$, we know that there is some choice of scalars $a_1, \dots, a_n \in \mathbf{F}$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Computing the numbers $a_1, \dots, a_n$ that satisfy the equation above can be difficult for an arbitrary basis of V. The next result shows, however, that this is easy for an orthonormal basis—just take $a_j = \langle v, e_j \rangle$.

6.30 Writing a vector as linear combination of orthonormal basis

Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

Now that we understand the usefulness of orthonormal bases, how do we go about finding them? For example, does $\mathcal{P}_m(\mathbf{R})$, with inner product given by integration on $[-1, 1]$? The next result will lead to answers to these questions.

The algorithm used in the next proof is called the ***Gram–Schmidt Procedure***. It gives a method for turning a linearly independent list into an orthonormal list with the same span as the original list.

6.31 Gram–Schmidt Procedure

Suppose $v_1, \dots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \dots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}.$$

Then $e_1, \dots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$$

for $j = 1, \dots, m$.

Now we can answer the question about the existence of orthonormal bases.

6.34 Existence of orthonormal basis

Every finite-dimensional inner product space has an orthonormal basis.

Sometimes we need to know not only that an orthonormal basis exists, but also that every orthonormal list can be extended to an orthonormal basis. In the next corollary, the Gram–Schmidt Procedure shows that such an extension is always possible.

6.35 Orthonormal list extends to orthonormal basis

Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.

Recall that a matrix is called upper triangular if all entries below the diagonal equal 0. In other words, an upper-triangular matrix looks like this:

$$\begin{pmatrix} * & & * \\ & \ddots & \\ 0 & & * \end{pmatrix},$$

where the 0 in the matrix above indicates that all entries below the diagonal equal 0, and asterisks are used to denote entries on and above the diagonal.

In the last chapter we showed that if V is a finite-dimensional complex vector space, then for each operator on V there is a basis with respect to which the matrix of the operator is upper triangular (see 5.27). Now that we are dealing with inner product spaces, we would like to know whether there exists an *orthonormal* basis with respect to which we have an upper-triangular matrix.

The next result shows that the existence of a basis with respect to which T has an upper-triangular matrix implies the existence of an orthonormal basis with this property. This result is true on both real and complex vector spaces (although on a real vector space, the hypothesis holds only for some operators).

6.37 Upper-triangular matrix with respect to orthonormal basis

Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.

The next result is an important application of the result above.

6.38 Schur's Theorem

Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some orthonormal basis of V.

Linear Functionals on Inner Product Spaces

Because linear maps into the scalar field $\mathbf{F}$ play a special role, we defined a special name for them in Section 3.F. That definition is repeated below in case you skipped Section 3.F.

6.39 Definition ***linear functional***

A ***linear functional*** on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

If $u \in V$, then the map that sends v to $\langle v, u \rangle$ is a linear functional on V. The next result shows that every linear functional on V is of this form.

6.42 Riesz Representation Theorem

Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

$$\varphi(v) = \langle v, u \rangle$$

for every $v \in V$.

6.C Orthogonal Complements and Minimization Problems

Orthogonal Complements

6.45 Definition ***orthogonal complement, $U^\perp$***

If U is a subset of V, then the ***orthogonal complement*** of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}.$$

For example, if U is a line in $\mathbf{R}^3$, then $U^\perp$ is the plane containing the origin that is perpendicular to U. If U is a plane in $\mathbf{R}^3$, then $U^\perp$ is the line containing the origin that is perpendicular to U.

6.46 Basic properties of orthogonal complement

(a) If U is a subset of V, then $U^\perp$ is a subspace of V.

(b) $\{0\}^\perp = V$.

(c) $V^\perp = \{0\}$.

(d) If U is a subset of V, then $U \cap U^\perp \subset \{0\}$.

(e) If U and W are subsets of V and $U \subset W$, then $W^\perp \subset U^\perp$.

Recall that if U, W are subspaces of V, then V is the direct sum of U and W (written $V = U \oplus W$) if each element of V can be written in exactly one way as a vector in U plus a vector in W (see 1.40).

The next result shows that every finite-dimensional subspace of V leads to a natural direct sum decomposition of V.

6.47 Direct sum of a subspace and its orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^{\perp}.$$

Now we can see how to compute $\dim U^{\perp}$ from $\dim U$.

6.50 Dimension of the orthogonal complement

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^{\perp} = \dim V - \dim U.$$

The next result is an important consequence of 6.47.

6.51 The orthogonal complement of the orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$U = (U^{\perp})^{\perp}.$$

We now define an operator $\mathcal{P}_U$ for each finite-dimensional subspace of V.

6.53 Definition *orthogonal projection, P_U*

Suppose U is a finite-dimensional subspace of V. The ***orthogonal projection*** of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. Then $P_U v = u$.

The direct sum decomposition $V = U \oplus U^{\perp}$ given by 6.47 shows that each $v \in V$ can be uniquely written in the form $v = u + w$ with $u \in U$ and $w \in U^{\perp}$. Thus $P_U v$ is well defined.

6.55 Properties of the orthogonal projection P_U

Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

(a) $P_U \in \mathcal{L}(V)$;

(b) $P_U u = u$ for every $u \in U$;

(c) $P_U w = 0$ for every $w \in U^\perp$;

(d) range $P_U = U$;

(e) null $P_U = U^\perp$;

(f) $v - P_U v \in U^\perp$;

(g) $P_U{}^2 = P_U$;

(h) $\|P_U v\| \leq \|v\|$;

(i) for every orthonormal basis $e_1, \ldots, e_m$ of U,

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Minimization Problems

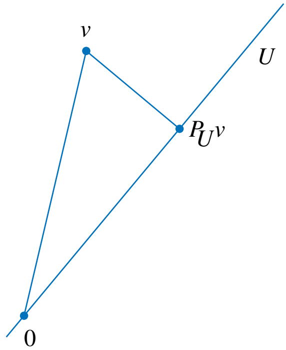
The following problem often arises: given a subspace U of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. The next proposition shows that this minimization problem is solved by taking $u = P_U v$.

6.56 Minimizing the distance to a subspace

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \leq \|v - u\|.$$

Furthermore, the inequality above is an equality if and only if $u = P_U v$.



$P_U v$ is the closest point in U to v.

The last result is often combined with the formula 6.55(i) to compute explicit solutions to minimization problems.

CHAPTER 7

Isaac Newton (1642–1727), as envisioned by British poet and artist William Blake in this 1795 painting.

Operators on Inner Product Spaces

The deepest results related to inner product spaces deal with the subject to which we now turn—operators on inner product spaces. By exploiting properties of the adjoint, we will develop a detailed description of several important classes of operators on inner product spaces.

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

7.1 Notation $\mathbf{F}$, V

- $\mathbf{F}$ denotes $\mathbf{R}$ or $\mathbf{C}$.
- V and W denote finite-dimensional inner product spaces over $\mathbf{F}$.

LEARNING OBJECTIVES FOR THIS CHAPTER

- adjoint
- Spectral Theorem
- positive operators
- isometries
- Polar Decomposition
- Singular Value Decomposition

7.A *Self-Adjoint and Normal Operators*

Adjoints

7.2 Definition ***adjoint,*** T^*

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

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

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

To see why the definition above makes sense, suppose $T \in \mathcal{L}(V, W)$. Fix $w \in W$. Consider the linear functional on V that maps $v \in V$ to $\langle Tv, w \rangle$; this linear functional depends on T and w. By the Riesz Representation Theorem (6.42), there exists a unique vector in V such that this linear functional is given by taking the inner product with it. We call this unique vector T^*w. In other words, T^*w is the unique vector in V such that $\langle Tv, w \rangle = \langle v, T^*w \rangle$ for every $v \in V$.

The proofs of the next two results use a common technique: flip T^* from one side of an inner product to become T on the other side.

7.5 The adjoint is a linear map

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

7.6 Properties of the adjoint

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

(b) $(\lambda T)^* = \bar{\lambda} T^*$ for all $\lambda \in \mathbf{F}$ and $T \in \mathcal{L}(V, W)$;

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

(d) $I^* = I$, where I is the identity operator on V;

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

The next result shows the relationship between the null space and the range of a linear map and its adjoint. The symbol $\Longleftrightarrow$ used in the proof means "if and only if"; this symbol could also be read to mean "is equivalent to".

7.7 Null space and range of T^*

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

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

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

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

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

7.8 Definition ***conjugate transpose***

The ***conjugate transpose*** of an m-by-n matrix is the n-by-m matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry.

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

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

7.10 The matrix of T^*

Let $T \in \mathcal{L}(V, W)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $f_1, \dots, f_m$ is an orthonormal basis of W. Then

$$\mathcal{M}\big(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n)\big)$$

is the conjugate transpose of

$$\mathcal{M}\big(T, (e_1, \dots, e_n), (f_1, \dots, f_m)\big).$$

Self-Adjoint Operators

Now we switch our attention to operators on inner product spaces. Thus instead of considering linear maps from V to W, we will be focusing on linear maps from V to V; recall that such linear maps are called operators.

7.11 Definition ***self-adjoint***

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

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

for all $v, w \in V$.

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

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

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

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

7.13 Eigenvalues of self-adjoint operators are real

Every eigenvalue of a self-adjoint operator is real.

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

7.14 Over $\mathbf{C}$, Tv is orthogonal to v for all v only for the 0 operator

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

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

for all $v \in V$. Then $T = 0$.

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.

Linear Algebra Abridged is generated from *Linear Algebra Done Right* (by Sheldon Axler, third edition) by excluding all proofs, examples, and exercises, along with most comments. The full version of *Linear Algebra Done Right* is available at springer.com and amazon.com in both printed and electronic forms.

7.15 Over $\mathbf{C}$, $\langle Tv, v\rangle$ is real for all v only for self-adjoint operators

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

$$\langle Tv, v\rangle \in \mathbf{R}$$

for every $v \in V$.

On a real inner product space V, a nonzero operator T might satisfy $\langle Tv, v\rangle = 0$ for all $v \in V$. However, the next result shows that this cannot happen for a self-adjoint operator.

7.16 If $T = T^*$ and $\langle Tv, v\rangle = 0$ for all v, then $T = 0$

Suppose T is a self-adjoint operator on V such that

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

for all $v \in V$. Then $T = 0$.

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 $T^* = T$.

In the next section we will see why normal operators are worthy of special attention.

The next result provides a simple characterization of normal operators.

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

An operator $T \in \mathcal{L}(V)$ is normal if and only if

$$\|Tv\| = \|T^*v\|$$

for all $v \in V$.

7.21 For T normal, T and T^* have the same eigenvectors

Suppose $T \in \mathcal{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T^* with eigenvalue $\bar{\lambda}$.

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.

7.B *The Spectral Theorem*

Recall that a diagonal matrix is a square matrix that is 0 everywhere except possibly along the diagonal. Recall also that an operator on V has a diagonal matrix with respect to a basis if and only if the basis consists of eigenvectors of the operator (see 5.41).

The nicest operators on V are those for which there is an *orthonormal* basis of V with respect to which the operator has a diagonal matrix. These are precisely the operators $T \in \mathcal{L}(V)$ such that there is an orthonormal basis of V consisting of eigenvectors of T. Our goal in this section is to prove the Spectral Theorem, which characterizes these operators as the normal operators when $\mathbf{F} = \mathbf{C}$ and as the self-adjoint operators when $\mathbf{F} = \mathbf{R}$. The Spectral Theorem is probably the most useful tool in the study of operators on inner product spaces.

Because the conclusion of the Spectral Theorem depends on $\mathbf{F}$, we will break the Spectral Theorem into two pieces, called the Complex Spectral Theorem and the Real Spectral Theorem. As is often the case in linear algebra, complex vector spaces are easier to deal with than real vector spaces. Thus we present the Complex Spectral Theorem first.

The Complex Spectral Theorem

The key part of the Complex Spectral Theorem (7.24) states that if $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$ is normal, then T has a diagonal matrix with respect to some orthonormal basis of V.

In the next result, the equivalence of (b) and (c) is easy (see 5.41). Thus we prove only that (c) implies (a) and that (a) implies (c).

7.24 Complex Spectral Theorem

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

(a) T is normal.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

The Real Spectral Theorem

We will need a few preliminary results, which apply to both real and complex inner product spaces, for our proof of the Real Spectral Theorem.

You could guess that the next result is true and even discover its proof by thinking about quadratic polynomials with real coefficients. Specifically, suppose $b, c \in \mathbf{R}$ and $b^2 < 4c$. Let x be a real number. Then

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0.$$

In particular, $x^2 + bx + c$ is an invertible real number (a convoluted way of saying that it is not 0). Replacing the real number x with a self-adjoint operator (recall the analogy between real numbers and self-adjoint operators), we are led to the result below.

7.26 Invertible quadratic expressions

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

$$T^2 + bT + cI$$

is invertible.

We know that every operator, self-adjoint or not, on a finite-dimensional nonzero complex vector space has an eigenvalue (see 5.21). Thus the next result tells us something new only for real inner product spaces.

7.27 Self-adjoint operators have eigenvalues

Suppose $V \neq \{0\}$ and $T \in \mathcal{L}(V)$ is a self-adjoint operator. Then T has an eigenvalue.

The next result shows that if U is a subspace of V that is invariant under a self-adjoint operator T, then $U^\perp$ is also invariant under T. Later we will show that the hypothesis that T is self-adjoint can be replaced with the weaker hypothesis that T is normal (see 9.30).

7.28 Self-adjoint operators and invariant subspaces

Suppose $T \in \mathcal{L}(V)$ is self-adjoint and U is a subspace of V that is invariant under T. Then

(a) $U^\perp$ is invariant under T;

(b) $T|_U \in \mathcal{L}(U)$ is self-adjoint;

(c) $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.

We can now prove the next result, which is one of the major theorems in linear algebra.

7.29 Real Spectral Theorem

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

(a) T is self-adjoint.

(b) V has an orthonormal basis consisting of eigenvectors of T.

(c) T has a diagonal matrix with respect to some orthonormal basis of V.

If $\mathbf{F} = \mathbf{C}$, then the Complex Spectral Theorem gives a complete description of the normal operators on V. A complete description of the self-adjoint operators on V then easily follows (they are the normal operators on V whose eigenvalues all are real).

If $\mathbf{F} = \mathbf{R}$, then the Real Spectral Theorem gives a complete description of the self-adjoint operators on V. In Chapter 9, we will give a complete description of the normal operators on V (see 9.34).

7.C *Positive Operators and Isometries*

Positive Operators

7.31 Definition ***positive operator***

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

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

for all $v \in V$.

If V is a complex vector space, then the requirement that T is self-adjoint can be dropped from the definition above (by 7.15).

7.33 Definition ***square root***

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

The characterizations of the positive operators in the next result correspond to characterizations of the nonnegative numbers among $\mathbf{C}$. Specifically, a complex number z is nonnegative if and only if it has a nonnegative square root, corresponding to condition (c). Also, z is nonnegative if and only if it has a real square root, corresponding to condition (d). Finally, z is nonnegative if and only if there exists a complex number w such that $z = \bar{w}w$, corresponding to condition (e).

7.35 Characterization of positive operators

Let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is positive;

(b) T is self-adjoint and all the eigenvalues of T are nonnegative;

(c) T has a positive square root;

(d) T has a self-adjoint square root;

(e) there exists an operator $R \in \mathcal{L}(V)$ such that $T = R^*R$.

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

7.36 Each positive operator has only one positive square root

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

Isometries

Operators that preserve norms are sufficiently important to deserve a name:

7.37 Definition ***isometry***

- An operator $S \in \mathcal{L}(V)$ is called an ***isometry*** if
$$\|Sv\| = \|v\|$$
for all $v \in V$.
- In other words, an operator is an isometry if it preserves norms.

For example, λI is an isometry whenever $\lambda \in \mathbf{F}$ satisfies $|\lambda| = 1$.

The next result provides several conditions that are equivalent to being an isometry. The equivalence of (a) and (b) shows that an operator is an isometry if and only if it preserves inner products. The equivalence of (a) and (c) [or (d)] shows that an operator is an isometry if and only if the list of columns of its matrix with respect to every [or some] basis is orthonormal.

7.42 Characterization of isometries

Suppose $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry;

(b) $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$;

(c) $Se_1, \dots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \dots, e_n$ in V;

(d) there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $Se_1, \dots, Se_n$ is orthonormal;

(e) $S^*S = I$;

(f) $SS^* = I$;

(g) S^* is an isometry;

(h) S is invertible and $S^{-1} = S^*$.

The previous result shows that every isometry is normal [see (a), (e), and (f) of 7.42]. Thus characterizations of normal operators can be used to give descriptions of isometries. We do this in the next result in the complex case and in Chapter 9 in the real case (see 9.36).

7.43 Description of isometries when $\mathbf{F} = \mathbf{C}$

Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry.

(b) There is an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value 1.

7.D *Polar Decomposition and Singular Value Decomposition*

Polar Decomposition

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

Another distinguished subset of $\mathbf{C}$ is the unit circle, which consists of the complex numbers z such that $|z| = 1$. The condition $|z| = 1$ is equivalent to the condition $\bar{z}z = 1$. Under our analogy, this would correspond to the condition $T^*T = I$, which is equivalent to T being an isometry (see 7.42). In other words, the unit circle in $\mathbf{C}$ corresponds to the isometries.

Continuing with our analogy, note that each complex number z except 0 can be written in the form

$$z = \left(\frac{z}{|z|}\right)|z| = \left(\frac{z}{|z|}\right)\sqrt{\bar{z}z},$$

where the first factor, namely, $z/|z|$, is an element of the unit circle. Our analogy leads us to guess that each operator $T \in \mathcal{L}(V)$ can be written as an isometry times $\sqrt{T^*T}$. That guess is indeed correct, as we now prove after defining the obvious notation, which is justified by 7.36.

7.44 Notation $\sqrt{T}$

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

Now we can state and prove the Polar Decomposition, which gives a beautiful description of an arbitrary operator on V. Note that T^*T is a positive operator for every $T \in \mathcal{L}(V)$, and thus $\sqrt{T^*T}$ is well defined.

7.45 Polar Decomposition

Suppose $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that

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

The Polar Decomposition (7.45) states that each operator on V is the product of an isometry and a positive operator. Thus we can write each operator on V as the product of two operators, each of which comes from a class that we can completely describe and that we understand reasonably well. The isometries are described by 7.43 and 9.36; the positive operators are described by the Spectral Theorem (7.24 and 7.29).

Specifically, consider the case $\mathbf{F} = \mathbf{C}$, and suppose $T = S\sqrt{T^*T}$ is a Polar Decomposition of an operator $T \in \mathcal{L}(V)$, where S is an isometry. Then there is an orthonormal basis of V with respect to which S has a diagonal matrix, and there is an orthonormal basis of V with respect to which $\sqrt{T^*T}$ has a diagonal matrix. **Warning:** there may not exist an orthonormal basis that simultaneously puts the matrices of both S and $\sqrt{T^*T}$ into these nice diagonal forms. In other words, S may require one orthonormal basis and $\sqrt{T^*T}$ may require a different orthonormal basis.

Singular Value Decomposition

The eigenvalues of an operator tell us something about the behavior of the operator. Another collection of numbers, called the singular values, is also useful. Recall that eigenspaces and the notation E are defined in 5.36.

7.49 Definition ***singular values***

Suppose $T \in \mathcal{L}(V)$. The ***singular values*** of T are the eigenvalues of $\sqrt{T^*T}$, with each eigenvalue λ repeated $\dim E(\lambda, \sqrt{T^*T})$ times.

The singular values of T are all nonnegative, because they are the eigenvalues of the positive operator $\sqrt{T^*T}$.

Each $T \in \mathcal{L}(V)$ has $\dim V$ singular values, as can be seen by applying the Spectral Theorem and 5.41 [see especially part (e)] to the positive (hence self-adjoint) operator $\sqrt{T^*T}$.

The next result shows that every operator on V has a clean description in terms of its singular values and two orthonormal bases of V.

7.51 Singular Value Decomposition

Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \dots, s_n$. Then there exist orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ of V such that

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

for every $v \in V$.

The Singular Value Decomposition allows us a rare opportunity to make good use of two different bases for the matrix of an operator. To do this, suppose $T \in \mathcal{L}(V)$. Let $s_1, \dots, s_n$ denote the singular values of T, and let $e_1, \dots, e_n$ and $f_1, \dots, f_n$ be orthonormal bases of V such that the Singular Value Decomposition 7.51 holds. Because $Te_j = s_j f_j$ for each j, we have

$$\mathcal{M}\big(T, (e_1, \dots, e_n), (f_1, \dots, f_n)\big) = \begin{pmatrix} s_1 & & 0 \\ & \ddots & \\ 0 & & s_n \end{pmatrix}.$$

In other words, every operator on V has a diagonal matrix with respect to some orthonormal bases of V, provided that we are permitted to use two different bases rather than a single basis as customary when working with operators.

To compute numeric approximations to the singular values of an operator T, first compute T^*T and then compute approximations to the eigenvalues of T^*T (good techniques exist for approximating eigenvalues of positive operators). The nonnegative square roots of these (approximate) eigenvalues of T^*T will be the (approximate) singular values of T. In other words, the singular values of T can be approximated without computing the square root of T^*T. The next result helps justify working with T^*T instead of $\sqrt{T^*T}$.

7.52 Singular values without taking square root of an operator

Suppose $T \in \mathcal{L}(V)$. Then the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, with each eigenvalue λ repeated $\dim E(\lambda, T^*T)$ times.

CHAPTER

8

Hypatia, the 5th century Egyptian mathematician and philosopher, as envisioned around 1900 by Alfred Seifert.

Operators on Complex Vector Spaces

In this chapter we delve deeper into the structure of operators, with most of the attention on complex vector spaces. An inner product does not help with this material, so we return to the general setting of a finite-dimensional vector space. To avoid some trivialities, we will assume that $V \neq \{0\}$. Thus our assumptions for this chapter are as follows:

8.1 **Notation** $\mathbf{F}$, V

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

LEARNING OBJECTIVES FOR THIS CHAPTER

- generalized eigenvectors and generalized eigenspaces
- characteristic polynomial and the Cayley–Hamilton Theorem
- decomposition of an operator
- minimal polynomial
- Jordan Form

8.A Generalized Eigenvectors and Nilpotent Operators

Null Spaces of Powers of an Operator

We begin this chapter with a study of null spaces of powers of an operator.

8.2 Sequence of increasing null spaces

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

$$\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^k \subset \operatorname{null} T^{k+1} \subset \cdots .$$

The next result says that if two consecutive terms in this sequence of subspaces are equal, then all later terms in the sequence are equal.

8.3 Equality in the sequence of null spaces

Suppose $T \in \mathcal{L}(V)$. Suppose m is a nonnegative integer such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Then

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \operatorname{null} T^{m+3} = \cdots .$$

The proposition above raises the question of whether there exists a nonnegative integer m such that $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. The proposition below shows that this equality holds at least when m equals the dimension of the vector space on which T operates.

8.4 Null spaces stop growing

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \operatorname{null} T^{n+2} = \cdots .$$

Unfortunately, it is not true that $V = \operatorname{null} T \oplus \operatorname{range} T$ for each $T \in \mathcal{L}(V)$. However, the following result is a useful substitute.

8.5 V is the direct sum of $\operatorname{null} T^{\dim V}$ and $\operatorname{range} T^{\dim V}$

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n .$$

Generalized Eigenvectors

Unfortunately, some operators do not have enough eigenvectors to lead to a good description. Thus in this subsection we introduce the concept of generalized eigenvectors, which will play a major role in our description of the structure of an operator.

To understand why we need more than eigenvectors, let's examine the question of describing an operator by decomposing its domain into invariant subspaces. Fix $T \in \mathcal{L}(V)$. We seek to describe T by finding a "nice" direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m,$$

where each U_j is a subspace of V invariant under T. The simplest possible nonzero invariant subspaces are 1-dimensional. A decomposition as above where each U_j is a 1-dimensional subspace of V invariant under T is possible if and only if V has a basis consisting of eigenvectors of T (see 5.41). This happens if and only if V has an eigenspace decomposition

8.8 $$V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T),$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T (see 5.41).

The Spectral Theorem in the previous chapter shows that if V is an inner product space, then a decomposition of the form 8.8 holds for every normal operator if $\mathbf{F} = \mathbf{C}$ and for every self-adjoint operator if $\mathbf{F} = \mathbf{R}$ because operators of those types have enough eigenvectors to form a basis of V (see 7.24 and 7.29).

Sadly, a decomposition of the form 8.8 may not hold for more general operators, even on a complex vector space. Generalized eigenvectors and generalized eigenspaces, which we now introduce, will remedy this situation.

8.9 **Definition** ***generalized eigenvector***

Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. A vector $v \in V$ is called a ***generalized eigenvector*** of T corresponding to λ if $v \neq 0$ and

$$(T - \lambda I)^j v = 0$$

for some positive integer j.

Although j is allowed to be an arbitrary integer in the equation

$$(T - \lambda I)^j v = 0$$

in the definition of a generalized eigenvector, we will soon prove that every generalized eigenvector satisfies this equation with $j = \dim V$.

8.10 **Definition** ***generalized eigenspace,*** $G(\lambda, T)$

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. The ***generalized eigenspace*** of T corresponding to λ, denoted $G(\lambda, T)$, is defined to be the set of all generalized eigenvectors of T corresponding to λ, along with the 0 vector.

Because every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the definition of generalized eigenvector), each eigenspace is contained in the corresponding generalized eigenspace. In other words, if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then

$$E(\lambda, T) \subset G(\lambda, T).$$

The next result implies that if $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$, then $G(\lambda, T)$ is a subspace of V (because the null space of each linear map on V is a subspace of V).

8.11 **Description of generalized eigenspaces**

Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbf{F}$. Then $G(\lambda, T) = \operatorname{null}(T - \lambda I)^{\dim V}$.

We saw earlier (5.10) that eigenvectors corresponding to distinct eigenvalues are linearly independent. Now we prove a similar result for generalized eigenvectors.

8.13 **Linearly independent generalized eigenvectors**

Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.

Nilpotent Operators

8.16 **Definition** ***nilpotent***

An operator is called ***nilpotent*** if some power of it equals 0.

The next result shows that we never need to use a power higher than the dimension of the space.

8.18 Nilpotent operator raised to dimension of domain is 0

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $N^{\dim V} = 0$.

Given an operator T on V, we want to find a basis of V such that the matrix of T with respect to this basis is as simple as possible, meaning that the matrix contains many 0's.

The next result shows that if N is nilpotent, then we can choose a basis of V such that the matrix of N with respect to this basis has more than half of its entries equal to 0. Later in this chapter we will do even better.

8.19 Matrix of a nilpotent operator

Suppose N is a nilpotent operator on V. Then there is a basis of V with respect to which the matrix of N has the form

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix};$$

here all entries on and below the diagonal are 0's.

8.B *Decomposition of an Operator*

Description of Operators on Complex Vector Spaces

We saw earlier that the domain of an operator might not decompose into eigenspaces, even on a finite-dimensional complex vector space. In this section we will see that every operator on a finite-dimensional complex vector space has enough generalized eigenvectors to provide a decomposition.

8.20 The null space and range of $p(T)$ are invariant under T

Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbf{F})$. Then null $p(T)$ and range $p(T)$ are invariant under T.

The following major result shows that every operator on a complex vector space can be thought of as composed of pieces, each of which is a nilpotent operator plus a scalar multiple of the identity. Actually we have already done the hard work in our discussion of the generalized eigenspaces $G(\lambda, T)$, so at this point the proof is easy.

8.21 Description of operators on complex vector spaces

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(b) each $G(\lambda_j, T)$ is invariant under T;

(c) each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

As we know, an operator on a complex vector space may not have enough eigenvectors to form a basis of the domain. The next result shows that on a complex vector space there are enough generalized eigenvectors to do this.

8.23 A basis of generalized eigenvectors

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.

Multiplicity of an Eigenvalue

If V is a complex vector space and $T \in \mathcal{L}(V)$, then the decomposition of V provided by 8.21 can be a powerful tool. The dimensions of the subspaces involved in this decomposition are sufficiently important to get a name.

8.24 Definition ***multiplicity***

- Suppose $T \in \mathcal{L}(V)$. The ***multiplicity*** of an eigenvalue λ of T is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$.
- In other words, the multiplicity of an eigenvalue λ of T equals $\dim \operatorname{null}(T - \lambda I)^{\dim V}$.

The second bullet point above is justified by 8.11.

8.26 Sum of the multiplicities equals $\dim V$

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then the sum of the multiplicities of all the eigenvalues of T equals $\dim V$.

The terms ***algebraic multiplicity*** and ***geometric multiplicity*** are used in some books. In case you encounter this terminology, be aware that the algebraic multiplicity is the same as the multiplicity defined here and the geometric multiplicity is the dimension of the corresponding eigenspace. In other words, if $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T, then

$$\begin{aligned}\text{algebraic multiplicity of } \lambda &= \dim \operatorname{null}(T - \lambda I)^{\dim V} = \dim G(\lambda, T),\\ \text{geometric multiplicity of } \lambda &= \dim \operatorname{null}(T - \lambda I) = \dim E(\lambda, T).\end{aligned}$$

Note that as defined above, the algebraic multiplicity also has a geometric meaning as the dimension of a certain null space. The definition of multiplicity given here is cleaner than the traditional definition that involves determinants; 10.25 implies that these definitions are equivalent.

Block Diagonal Matrices

To interpret our results in matrix form, we make the following definition, generalizing the notion of a diagonal matrix.

If each matrix A_j in the definition below is a 1-by-1 matrix, then we actually have a diagonal matrix.

8.27 Definition ***block diagonal matrix***

A ***block diagonal matrix*** is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \dots, A_m$ are square matrices lying along the diagonal and all the other entries of the matrix equal 0.

Note that in the next result we get many more zeros in the matrix of T than are needed to make it upper triangular.

8.29 Block diagonal matrix with upper-triangular blocks

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T, with multiplicities $d_1, \dots, d_m$. Then there is a basis of V with respect to which T has a block diagonal matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_j is a d_j-by-d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}.$$

When we discuss the Jordan Form in Section 8.D, we will see that we can find a basis with respect to which an operator T has a matrix with even more 0's than promised by 8.29. However, 8.29 and its equivalent companion 8.21 are already quite powerful. For example, in the next subsection we will use 8.21 to show that every invertible operator on a complex vector space has a square root.

Square Roots

Recall that a square root of an operator $T \in \mathcal{L}(V)$ is an operator $R \in \mathcal{L}(V)$ such that $R^2 = T$ (see 7.33). Every complex number has a square root, but not every operator on a complex vector space has a square root.

8.31 Identity plus nilpotent has a square root

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.

The previous lemma is valid on real and complex vector spaces. However, the next result holds only on complex vector spaces. For example, the operator of multiplication by -1 on the 1-dimensional real vector space **R** has no square root.

8.33 Over C, invertible operators have square roots

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Then T has a square root.

By imitating the techniques in this section, you should be able to prove that if V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible, then T has a k^{th} root for every positive integer k.

8.C Characteristic and Minimal Polynomials

The Cayley–Hamilton Theorem

The next definition associates a polynomial with each operator on V if $\mathbf{F} = \mathbf{C}$. For $\mathbf{F} = \mathbf{R}$, the corresponding definition will be given in the next chapter.

8.34 Definition ***characteristic polynomial***

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T, with multiplicities $d_1, \dots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the ***characteristic polynomial*** of T.

8.36 Degree and zeros of characteristic polynomial

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

(a) the characteristic polynomial of T has degree $\dim V$;

(b) the zeros of the characteristic polynomial of T are the eigenvalues of T.

Most texts define the characteristic polynomial using determinants (the two definitions are equivalent by 10.25). The approach taken here, which is considerably simpler, leads to the following easy proof of the Cayley–Hamilton Theorem. In the next chapter, we will see that this result also holds on real vector spaces (see 9.24).

8.37 Cayley–Hamilton Theorem

Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

The Minimal Polynomial

In this subsection we introduce another important polynomial associated with each operator. We begin with the following definition.

8.38 Definition ***monic polynomial***

A ***monic polynomial*** is a polynomial whose highest-degree coefficient equals 1.

8.40 Minimal polynomial

Suppose $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$.

The last result justifies the following definition.

8.43 Definition ***minimal polynomial***

Suppose $T \in \mathcal{L}(V)$. Then the ***minimal polynomial*** of T is the unique monic polynomial p of smallest degree such that $p(T) = 0$.

The proof of the last result shows that the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$. The Cayley–Hamilton Theorem (8.37) tells us that if V is a complex vector space, then the minimal polynomial of each operator on V has degree at most $\dim V$. This remarkable improvement also holds on real vector spaces, as we will see in the next chapter.

Suppose you are given the matrix (with respect to some basis) of an operator $T \in \mathcal{L}(V)$. You could program a computer to find the minimal polynomial of T as follows: Consider the system of linear equations

8.44 $$a_0\mathcal{M}(I) + a_1\mathcal{M}(T) + \cdots + a_{m-1}\mathcal{M}(T)^{m-1} = -\mathcal{M}(T)^m$$

for successive values of $m = 1, 2, \ldots$ until this system of equations has a solution $a_0, a_1, a_2, \ldots, a_{m-1}$. The scalars $a_0, a_1, a_2, \ldots, a_{m-1}, 1$ will then be the coefficients of the minimal polynomial of T. All this can be computed using a familiar and fast (for a computer) process such as Gaussian elimination.

The next result completely characterizes the polynomials that when applied to an operator give the 0 operator.

8.46 $q(T) = 0$ implies q is a multiple of the minimal polynomial

Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.

The next result is stated only for complex vector spaces, because we have not yet defined the characteristic polynomial when $\mathbf{F} = \mathbf{R}$. However, the result also holds for real vector spaces, as we will see in the next chapter.

8.48 Characteristic polynomial is a multiple of minimal polynomial

Suppose $\mathbf{F} = \mathbf{C}$ and $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

We know (at least when $\mathbf{F} = \mathbf{C}$) that the zeros of the characteristic polynomial of T are the eigenvalues of T (see 8.36). Now we show that the minimal polynomial has the same zeros (although the multiplicities of these zeros may differ).

8.49 Eigenvalues are the zeros of the minimal polynomial

Let $T \in \mathcal{L}(V)$. Then the zeros of the minimal polynomial of T are precisely the eigenvalues of T.

8.D *Jordan Form*

We know that if V is a complex vector space, then for every $T \in \mathcal{L}(V)$ there is a basis of V with respect to which T has a nice upper-triangular matrix (see 8.29). In this section we will see that we can do even better—there is a basis of V with respect to which the matrix of T contains 0's everywhere except possibly on the diagonal and the line directly above the diagonal.

For the matrix interpretation of the next result, see the first part of the proof of 8.60.

8.55 Basis corresponding to a nilpotent operator

Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \dots, v_n \in V$ and nonnegative integers $m_1, \dots, m_n$ such that

(a) $N^{m_1} v_1, \dots, N v_1, v_1, \dots, N^{m_n} v_n, \dots, N v_n, v_n$ is a basis of V;

(b) $N^{m_1+1} v_1 = \cdots = N^{m_n+1} v_n = 0$.

In the next definition, the diagonal of each A_j is filled with some eigenvalue λ_j of T, the line directly above the diagonal of A_j is filled with 1's, and all other entries in A_j are 0 (to understand why each λ_j is an eigenvalue of T, see 5.32). The λ_j's need not be distinct. Also, A_j may be a 1-by-1 matrix (λ_j) containing just an eigenvalue of T.

8.59 Definition *Jordan basis*

Suppose $T \in \mathcal{L}(V)$. A basis of V is called a ***Jordan basis*** for T if with respect to this basis T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

where each A_j is an upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}.$$

8.60 Jordan Form

Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.

CHAPTER

9

Euclid explaining geometry (from The School of Athens, *painted by Raphael around 1510).*

Operators on Real Vector Spaces

In the last chapter we learned about the structure of an operator on a finite-dimensional complex vector space. In this chapter, we will use our results about operators on complex vector spaces to learn about operators on real vector spaces.

Our assumptions for this chapter are as follows:

9.1 Notation $\mathbf{F}$, V

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

LEARNING OBJECTIVES FOR THIS CHAPTER

- complexification of a real vector space
- complexification of an operator on a real vector space
- operators on finite-dimensional real vector spaces have an eigenvalue or a 2-dimensional invariant subspace
- characteristic polynomial and the Cayley–Hamilton Theorem
- description of normal operators on a real inner product space
- description of isometries on a real inner product space

9.A Complexification

Complexification of a Vector Space

As we will soon see, a real vector space V can be embedded, in a natural way, in a complex vector space called the complexification of V. Each operator on V can be extended to an operator on the complexification of V. Our results about operators on complex vector spaces can then be translated to information about operators on real vector spaces.

We begin by defining the complexification of a real vector space.

9.2 Definition ***complexification of*** V, $V_{\mathbf{C}}$

Suppose V is a real vector space.

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

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

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

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

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

Motivation for the definition above of complex scalar multiplication comes from usual algebraic properties and the identity $i^2 = -1$. If you remember the motivation, then you do not need to memorize the definition above.

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

9.3 $V_{\mathbf{C}}$ **is a complex vector space.**

Suppose V is a real vector space. Then with the definitions of addition and scalar multiplication as above, $V_{\mathbf{C}}$ is a complex vector space.

Note that the additive identity of $V_{\mathbf{C}}$ is $0 + i0$, which we write as just 0.

Probably everything that you think should work concerning complexification does work, usually with a straightforward verification, as illustrated by the next result.

9.4 Basis of V is basis of $V_{\mathbf{C}}$

Suppose V is a real vector space.

(a) If $v_1, \dots, v_n$ is a basis of V (as a real vector space), then $v_1, \dots, v_n$ is a basis of $V_{\mathbf{C}}$ (as a complex vector space).

(b) The dimension of $V_{\mathbf{C}}$ (as a complex vector space) equals the dimension of V (as a real vector space).

Complexification of an Operator

Now we can define the complexification of an operator.

9.5 Definition ***complexification of T, $T_{\mathbf{C}}$***

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. The ***complexification*** of T, denoted $T_{\mathbf{C}}$, is the operator $T_{\mathbf{C}} \in \mathcal{L}(V_{\mathbf{C}})$ defined by

$$T_{\mathbf{C}}(u + iv) = Tu + iTv$$

for $u, v \in V$.

You should verify that if V is a real vector space and $T \in \mathcal{L}(V)$, then $T_{\mathbf{C}}$ is indeed in $\mathcal{L}(V_{\mathbf{C}})$. The key point here is that our definition of complex scalar multiplication can be used to show that $T_{\mathbf{C}}\big(\lambda(u + iv)\big) = \lambda T_{\mathbf{C}}(u + iv)$ for all $u, v \in V$ and all **complex** numbers λ.

The next result makes sense because 9.4 tells us that a basis of a real vector space is also a basis of its complexification. The proof of the next result follows immediately from the definitions.

9.7 Matrix of $T_{\mathbf{C}}$ equals matrix of T

Suppose V is a real vector space with basis $v_1, \dots, v_n$ and $T \in \mathcal{L}(V)$. Then $\mathcal{M}(T) = \mathcal{M}(T_{\mathbf{C}})$, where both matrices are with respect to the basis $v_1, \dots, v_n$.

Complexification of an operator could have been defined using matrices, but the approach taken here is more natural because it does not depend on the choice of a basis.

We know that every operator on a nonzero finite-dimensional complex vector space has an eigenvalue (see 5.21) and thus has a 1-dimensional invariant subspace.

9.8 Every operator has an invariant subspace of dimension 1 or 2

Every operator on a nonzero finite-dimensional vector space has an invariant subspace of dimension 1 or 2.

The Minimal Polynomial of the Complexification

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Repeated application of the definition of $T_{\mathbf{C}}$ shows that

9.9
$$(T_{\mathbf{C}})^n(u + iv) = T^n u + iT^n v$$

for every positive integer n and all $u, v \in V$.

Notice that the next result implies that the minimal polynomial of $T_{\mathbf{C}}$ has real coefficients.

9.10 Minimal polynomial of $T_{\mathbf{C}}$ equals minimal polynomial of T

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the minimal polynomial of $T_{\mathbf{C}}$ equals the minimal polynomial of T.

Eigenvalues of the Complexification

Now we turn to questions about the eigenvalues of the complexification of an operator. Again, everything that we expect to work indeed works easily.

We begin with a result showing that the real eigenvalues of $T_{\mathbf{C}}$ are precisely the eigenvalues of T. We give two different proofs of this result. The first proof is more elementary, but the second proof is shorter and gives some useful insight.

9.11 Real eigenvalues of $T_{\mathbf{C}}$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{R}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if λ is an eigenvalue of T.

Our next result shows that $T_{\mathbf{C}}$ behaves symmetrically with respect to an eigenvalue λ and its complex conjugate $\bar{\lambda}$.

9.12 $T_{\mathbf{C}} - \lambda I$ and $T_{\mathbf{C}} - \bar{\lambda} I$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, $\lambda \in \mathbf{C}$, j is a nonnegative integer, and $u, v \in V$. Then

$$(T_{\mathbf{C}} - \lambda I)^j (u + iv) = 0 \quad \text{if and only if} \quad (T_{\mathbf{C}} - \bar{\lambda} I)^j (u - iv) = 0.$$

An important consequence of the result above is the next result, which states that if a number is an eigenvalue of $T_{\mathbf{C}}$, then its complex conjugate is also an eigenvalue of $T_{\mathbf{C}}$.

9.16 Nonreal eigenvalues of $T_{\mathbf{C}}$ come in pairs

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$. Then λ is an eigenvalue of $T_{\mathbf{C}}$ if and only if $\bar{\lambda}$ is an eigenvalue of $T_{\mathbf{C}}$.

By definition, the eigenvalues of an operator on a real vector space are real numbers. Thus when mathematicians sometimes informally mention the complex eigenvalues of an operator on a real vector space, what they have in mind is the eigenvalues of the complexification of the operator.

Recall that the multiplicity of an eigenvalue is defined to be the dimension of the generalized eigenspace corresponding to that eigenvalue (see 8.24). The next result states that the multiplicity of an eigenvalue of a complexification equals the multiplicity of its complex conjugate.

9.17 Multiplicity of λ equals multiplicity of $\bar{\lambda}$

Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbf{C}$ is an eigenvalue of $T_{\mathbf{C}}$. Then the multiplicity of λ as an eigenvalue of $T_{\mathbf{C}}$ equals the multiplicity of $\bar{\lambda}$ as an eigenvalue of $T_{\mathbf{C}}$.

9.19 Operator on odd-dimensional vector space has eigenvalue

Every operator on an odd-dimensional real vector space has an eigenvalue.

Characteristic Polynomial of the Complexification

In the previous chapter we defined the characteristic polynomial of an operator on a finite-dimensional complex vector space (see 8.34). The next result is a key step toward defining the characteristic polynomial for operators on finite-dimensional real vector spaces.

9.20 Characteristic polynomial of $T_{\mathbf{C}}$

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the coefficients of the characteristic polynomial of $T_{\mathbf{C}}$ are all real.

Now we can define the characteristic polynomial of an operator on a finite-dimensional real vector space to be the characteristic polynomial of its complexification.

9.21 Definition ***Characteristic polynomial***

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the ***characteristic polynomial*** of T is defined to be the characteristic polynomial of $T_{\mathbf{C}}$.

In the next result, the eigenvalues of T are all real (because T is an operator on a real vector space).

9.23 Degree and zeros of characteristic polynomial

Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then

(a) the coefficients of the characteristic polynomial of T are all real;

(b) the characteristic polynomial of T has degree $\dim V$;

(c) the eigenvalues of T are precisely the real zeros of the characteristic polynomial of T.

In the previous chapter, we proved the Cayley–Hamilton Theorem (8.37) for complex vector spaces. Now we can also prove it for real vector spaces.

9.24 Cayley–Hamilton Theorem

Suppose $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.

We can now prove another result that we previously knew only in the complex case.

9.26 Characteristic polynomial is a multiple of minimal polynomial

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

(a) the degree of the minimal polynomial of T is at most $\dim V$;

(b) the characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.

9.B *Operators on Real Inner Product Spaces*

We now switch our focus to the context of inner product spaces. We will give a complete description of normal operators on real inner product spaces; a key step in the proof of this result (9.34) requires the result from the previous section that an operator on a finite-dimensional real vector space has an invariant subspace of dimension 1 or 2 (9.8).

Normal Operators on Real Inner Product Spaces

The Complex Spectral Theorem (7.24) gives a complete description of normal operators on complex inner product spaces. In this subsection we will give a complete description of normal operators on real inner product spaces.

9.27 Normal but not self-adjoint operators

Suppose V is a 2-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal but not self-adjoint.

(b) The matrix of T with respect to every orthonormal basis of V has the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b \neq 0$.

(c) The matrix of T with respect to some orthonormal basis of V has the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b > 0$.

The next result tells us that a normal operator restricted to an invariant subspace is normal. This will allow us to use induction on $\dim V$ when we prove our description of normal operators (9.34).

9.30 Normal operators and invariant subspaces

Suppose V is an inner product space, $T \in \mathcal{L}(V)$ is normal, and U is a subspace of V that is invariant under T. Then

(a) $U^{\perp}$ is invariant under T;

(b) U is invariant under T^*;

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

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

Our next result shows that normal operators on real inner product spaces come close to having diagonal matrices. Specifically, we get block diagonal matrices, with each block having size at most 2-by-2.

We cannot expect to do better than the next result, because on a real inner product space there exist normal operators that do not have a diagonal matrix with respect to any basis. For example, the operator $T \in \mathcal{L}(\mathbf{R}^2)$ defined by $T(x, y) = (-y, x)$ is normal (as you should verify) but has no eigenvalues; thus this particular T does not have even an upper-triangular matrix with respect to any basis of $\mathbf{R}^2$.

9.34 Characterization of normal operators when $\mathbf{F} = \mathbf{R}$

Suppose V is a real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:

(a) T is normal.

(b) There is an orthonormal basis of V with respect to which T has a block diagonal matrix such that each block is a 1-by-1 matrix or a 2-by-2 matrix of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

with $b > 0$.

Isometries on Real Inner Product Spaces

The next result shows that every isometry on a real inner product space is composed of pieces that are rotations on 2-dimensional subspaces, pieces that equal the identity operator, and pieces that equal multiplication by -1.

9.36 Description of isometries when $\mathbf{F} = \mathbf{R}$

Suppose V is a real inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent:

(a) S is an isometry.

(b) There is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block on the diagonal is a 1-by-1 matrix containing 1 or -1 or is a 2-by-2 matrix of the form

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

with $\theta \in (0, \pi)$.

CHAPTER
10

British mathematician and pioneer computer scientist Ada Lovelace (1815–1852), as painted by Alfred Chalon in this 1840 portrait.

Trace and Determinant

Throughout this book our emphasis has been on linear maps and operators rather than on matrices. In this chapter we pay more attention to matrices as we define the trace and determinant of an operator and then connect these notions to the corresponding notions for matrices. The book concludes with an explanation of the important role played by determinants in the theory of volume and integration.

Our assumptions for this chapter are as follows:

10.1 **Notation** $\mathbf{F}$, V

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

LEARNING OBJECTIVES FOR THIS CHAPTER

- change of basis and its effect upon the matrix of an operator
- trace of an operator and of a matrix
- determinant of an operator and of a matrix
- determinants and volume

10.A *Trace*

For our study of the trace and determinant, we will need to know how the matrix of an operator changes with a change of basis. Thus we begin this chapter by developing the necessary material about change of basis.

Change of Basis

With respect to every basis of V, the matrix of the identity operator $I \in \mathcal{L}(V)$ is the diagonal matrix with 1's on the diagonal and 0's elsewhere. We also use the symbol I for the name of this matrix, as shown in the next definition.

10.2 Definition ***identity matrix, I***

Suppose n is a positive integer. The n-by-n diagonal matrix

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

is called the ***identity matrix*** and is denoted I.

Note that we use the symbol I to denote the identity operator (on all vector spaces) and the identity matrix (of all possible sizes). You should always be able to tell from the context which particular meaning of I is intended. For example, consider the equation $\mathcal{M}(I) = I$; on the left side I denotes the identity operator, and on the right side I denotes the identity matrix.

If A is a square matrix (with entries in $\mathbf{F}$, as usual) with the same size as I, then $AI = IA = A$, as you should verify.

10.3 Definition ***invertible, inverse,*** A^{-1}

A square matrix A is called ***invertible*** if there is a square matrix B of the same size such that $AB = BA = I$; we call B the ***inverse*** of A and denote it by A^{-1}.

The same proof as used in 3.54 shows that if A is an invertible square matrix, then there is a unique matrix B such that $AB = BA = I$ (and thus the notation $B = A^{-1}$ is justified).

In Section 3.C we defined the matrix of a linear map from one vector space to another with respect to two bases—one basis of the first vector space and another basis of the second vector space. When we study operators, which are linear maps from a vector space to itself, we almost always use the same basis for both vector spaces (after all, the two vector spaces in question are equal). Thus we usually refer to the matrix of an operator with respect to a basis and display at most one basis because we are using one basis in two capacities.

The next result is one of the unusual cases in which we use two different bases even though we have operators from a vector space to itself. It is just a convenient restatement of 3.43 (with U and W both equal to V), but now we are being more careful to include the various bases explicitly in the notation. The result below holds because we defined matrix multiplication to make it true—see 3.43 and the material preceding it.

10.4 The matrix of the product of linear maps

Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ and $w_1, \dots, w_n$ are all bases of V. Suppose $S, T \in \mathcal{L}(V)$. Then

$$\mathcal{M}\big(ST, (u_1, \dots, u_n), (w_1, \dots, w_n)\big) = \\ \mathcal{M}\big(S, (v_1, \dots, v_n), (w_1, \dots, w_n)\big)\mathcal{M}\big(T, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

The next result deals with the matrix of the identity operator I with respect to two different bases. Note that the k^{th} column of the matrix $\mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$ consists of the scalars needed to write u_k as a linear combination of $v_1, \dots, v_n$.

10.5 Matrix of the identity with respect to two bases

Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then the matrices $\mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$ and $\mathcal{M}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big)$ are invertible, and each is the inverse of the other.

Now we can see how the matrix of T changes when we change bases. In the result below, we have two different bases of V. Recall that the notation $\mathcal{M}\big(T, (u_1, \dots, u_n)\big)$ is shorthand for $\mathcal{M}\big(T, (u_1, \dots, u_n), (u_1, \dots, u_n)\big)$

10.7 Change of basis formula

Suppose $T \in \mathcal{L}(V)$. Let $u_1, \dots, u_n$ and $v_1, \dots, v_n$ be bases of V. Let $A = \mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big)$. Then

$$\mathcal{M}\big(T, (u_1, \dots, u_n)\big) = A^{-1}\mathcal{M}\big(T, (v_1, \dots, v_n)\big)A.$$

Trace: A Connection Between Operators and Matrices

Suppose $T \in \mathcal{L}(V)$ and λ is an eigenvalue of T. Let $n = \dim V$. Recall that we defined the multiplicity of λ to be the dimension of the generalized eigenspace $G(\lambda, T)$ (see 8.24) and that this multiplicity equals $\dim \operatorname{null}(T - \lambda I)^n$ (see 8.11). Recall also that if V is a complex vector space, then the sum of the multiplicities of all the eigenvalues of T equals n (see 8.26).

In the definition below, the sum of the eigenvalues "with each eigenvalue repeated according to its multiplicity" means that if $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with multiplicities $d_1, \dots, d_m$, then the sum is

$$d_1\lambda_1 + \cdots + d_m\lambda_m.$$

Or if you prefer to list the eigenvalues with each repeated according to its multiplicity, then the eigenvalues could be denoted $\lambda_1, \dots, \lambda_n$ (where the index n equals $\dim V$) and the sum is

$$\lambda_1 + \cdots + \lambda_n.$$

10.9 Definition ***trace of an operator***

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

- If $\mathbf{F} = \mathbf{C}$, then the ***trace*** of T is the sum of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the ***trace*** of T is the sum of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The trace of T is denoted by $\operatorname{trace} T$.

The trace has a close connection with the characteristic polynomial. Suppose $\lambda_1, \dots, \lambda_n$ are the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with each eigenvalue repeated according to its multiplicity. Then by definition (see 8.34 and 9.21), the characteristic polynomial of T equals

$$(z - \lambda_1) \cdots (z - \lambda_n).$$

Expanding the polynomial above, we can write the characteristic polynomial of T in the form

10.11 $$z^n - (\lambda_1 + \cdots + \lambda_n)z^{n-1} + \cdots + (-1)^n(\lambda_1 \cdots \lambda_n).$$

The expression above immediately leads to the following result.

10.12 Trace and characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then trace T equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T.

Most of the rest of this section is devoted to discovering how to compute trace T from the matrix of T (with respect to an arbitrary basis).

10.13 Definition *trace of a matrix*

The ***trace*** of a square matrix A, denoted trace A, is defined to be the sum of the diagonal entries of A.

Now we have defined the trace of an operator and the trace of a square matrix, using the same word "trace" in two different contexts. This would be bad terminology unless the two concepts turn out to be essentially the same. As we will see, it is indeed true that $\operatorname{trace} T = \operatorname{trace} \mathcal{M}\big(T, (v_1, \dots, v_n)\big)$, where $v_1, \dots, v_n$ is an arbitrary basis of V. We will need the following result for the proof.

10.14 Trace of AB equals trace of BA

If A and B are square matrices of the same size, then

$$\operatorname{trace}(AB) = \operatorname{trace}(BA).$$

Now we can prove that the sum of the diagonal entries of the matrix of an operator is independent of the basis with respect to which the matrix is computed.

10.15 Trace of matrix of operator does not depend on basis

Let $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then

$$\operatorname{trace} \mathcal{M}\big(T, (u_1, \dots, u_n)\big) = \operatorname{trace} \mathcal{M}\big(T, (v_1, \dots, v_n)\big).$$

The result below, which is the most important result in this section, states that the trace of an operator equals the sum of the diagonal entries of the matrix of the operator. This theorem does not specify a basis because, by the result above, the sum of the diagonal entries of the matrix of an operator is the same for every choice of basis.

10.16 Trace of an operator equals trace of its matrix

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

We can use 10.16 to give easy proofs of some useful properties about traces of operators by shifting to the language of traces of matrices, where certain properties have already been proved or are obvious. The proof of the next result is an example of this technique. The eigenvalues of $S + T$ are not, in general, formed from adding together eigenvalues of S and eigenvalues of T. Thus the next result would be difficult to prove without using 10.16.

10.18 Trace is additive

Suppose $S, T \in \mathcal{L}(V)$. Then $\operatorname{trace}(S + T) = \operatorname{trace} S + \operatorname{trace} T$.

The techniques we have developed have the following curious consequence. A generalization of this result to infinite-dimensional vector spaces has important consequences in modern physics, particularly in quantum theory.

10.19 The identity is not the difference of ST and TS

There do not exist operators $S, T \in \mathcal{L}(V)$ such that $ST - TS = I$.

10.B *Determinant*

Determinant of an Operator

Now we are ready to define the determinant of an operator. Notice that the definition below mimics the approach we took when defining the trace, with the product of the eigenvalues replacing the sum of the eigenvalues.

10.20 Definition ***determinant of an operator,*** det T

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

- If $\mathbf{F} = \mathbf{C}$, then the ***determinant*** of T is the product of the eigenvalues of T, with each eigenvalue repeated according to its multiplicity.
- If $\mathbf{F} = \mathbf{R}$, then the ***determinant*** of T is the product of the eigenvalues of $T_{\mathbf{C}}$, with each eigenvalue repeated according to its multiplicity.

The determinant of T is denoted by det T.

If $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with multiplicities $d_1, \dots, d_m$, then the definition above implies

$$\det T = \lambda_1^{d_1} \cdots \lambda_m^{d_m}.$$

Or if you prefer to list the eigenvalues with each repeated according to its multiplicity, then the eigenvalues could be denoted $\lambda_1, \dots, \lambda_n$ (where the index n equals $\dim V$) and the definition above implies

$$\det T = \lambda_1 \cdots \lambda_n.$$

The determinant has a close connection with the characteristic polynomial. Suppose $\lambda_1, \dots, \lambda_n$ are the eigenvalues of T (or of $T_{\mathbf{C}}$ if V is a real vector space) with each eigenvalue repeated according to its multiplicity. Then the expression for the characteristic polynomial of T given by 10.11 gives the following result.

10.22 Determinant and characteristic polynomial

Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\det T$ equals $(-1)^n$ times the constant term of the characteristic polynomial of T.

Combining the result above and 10.12, we have the following result.

10.23 Characteristic polynomial, trace, and determinant

Suppose $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T can be written as

$$z^n - (\operatorname{trace} T)z^{n-1} + \cdots + (-1)^n(\det T).$$

We turn now to some simple but important properties of determinants. Later we will discover how to calculate $\det T$ from the matrix of T (with respect to an arbitrary basis).

The crucial result below has an easy proof due to our definition.

10.24 Invertible is equivalent to nonzero determinant

An operator on V is invertible if and only if its determinant is nonzero.

Some textbooks take the result below as the definition of the characteristic polynomial and then have our definition of the characteristic polynomial as a consequence.

10.25 Characteristic polynomial of T equals $\det(zI - T)$

Suppose $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T equals $\det(zI - T)$.

Determinant of a Matrix

Our next task is to discover how to compute $\det T$ from the matrix of T (with respect to an arbitrary basis). Let's start with the easiest situation. Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and we choose a basis of V as in 8.29. With respect to that basis, T has an upper-triangular matrix with the diagonal of the matrix containing precisely the eigenvalues of T, each repeated according to its multiplicity. Thus $\det T$ equals the product of the diagonal entries of $\mathcal{M}(T)$ with respect to that basis.

When dealing with the trace in the previous section, we discovered that the formula (trace = sum of diagonal entries) that worked for the upper-triangular matrix given by 8.29 also worked with respect to an arbitrary basis. Could that also work for determinants? In other words, is the determinant of an operator equal to the product of the diagonal entries of the matrix of the operator with respect to an arbitrary basis?

Unfortunately, the determinant is more complicated than the trace. In particular, $\det T$ need not equal the product of the diagonal entries of $\mathcal{M}(T)$ with respect to an arbitrary basis.

For each square matrix A, we want to define the determinant of A, denoted $\det A$, so that $\det T = \det \mathcal{M}(T)$ regardless of which basis is used to compute $\mathcal{M}(T)$.

10.27 Definition ***permutation,*** perm n

- A ***permutation*** of $(1, \dots, n)$ is a list $(m_1, \dots, m_n)$ that contains each of the numbers $1, \dots, n$ exactly once.
- The set of all permutations of $(1, \dots, n)$ is denoted perm n.

For example, $(2, 3, 4, 5, 1) \in \text{perm } 5$. You should think of an element of perm n as a rearrangement of the first n integers.

10.30 Definition ***sign of a permutation***

- The ***sign*** of a permutation $(m_1, \dots, m_n)$ is defined to be 1 if the number of pairs of integers (j, k) with $1 \leq j < k \leq n$ such that j appears after k in the list $(m_1, \dots, m_n)$ is even and -1 if the number of such pairs is odd.
- In other words, the sign of a permutation equals 1 if the natural order has been changed an even number of times and equals -1 if the natural order has been changed an odd number of times.

The next result shows that interchanging two entries of a permutation changes the sign of the permutation.

10.32 Interchanging two entries in a permutation

Interchanging two entries in a permutation multiplies the sign of the permutation by -1.

10.33 Definition ***determinant of a matrix,*** det A

Suppose A is an n-by-n matrix

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}.$$

The ***determinant*** of A, denoted det A, is defined by

$$\det A = \sum_{(m_1,\dots,m_n) \in \operatorname{perm} n} \big(\operatorname{sign}(m_1, \dots, m_n)\big) A_{m_1,1} \cdots A_{m_n,n}.$$

To make sure you understand this process, you should now find the formula for the determinant of an arbitrary 3-by-3 matrix using just the definition given above.

Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and we choose a basis of V as in 8.29. With respect to that basis, T has an upper-triangular matrix with the diagonal of the matrix containing precisely the eigenvalues of T, each repeated according to its multiplicity.

Our goal is to prove that $\det T = \det \mathcal{M}(T)$ for every basis of V, not just the basis from 8.29. To do this, we will need to develop some properties of determinants of matrices. The result below is the first of the properties we will need.

10.36 Interchanging two columns in a matrix

Suppose A is a square matrix and B is the matrix obtained from A by interchanging two columns. Then

$$\det A = -\det B.$$

If $T \in \mathcal{L}(V)$ and the matrix of T (with respect to some basis) has two equal columns, then T is not injective and hence $\det T = 0$. Although this comment makes the next result plausible, it cannot be used in the proof, because we do not yet know that $\det T = \det \mathcal{M}(T)$ for every choice of basis.

10.37 Matrices with two equal columns

If A is a square matrix that has two equal columns, then $\det A = 0$.

Recall from 3.44 that if A is an n-by-n matrix

$$A = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix},$$

then we can think of the k^{th} column of A as an n-by-1 matrix denoted $A_{\cdot,k}$:

$$A_{\cdot,k} = \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{n,k} \end{pmatrix}.$$

Note that $A_{j,k}$, with two subscripts, denotes an entry of A, whereas $A_{\cdot,k}$, with a dot as a placeholder and one subscript, denotes a column of A. This notation allows us to write A in the form

$$\begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,n} \end{pmatrix},$$

which will be useful.

The next result shows that a permutation of the columns of a matrix changes the determinant by a factor of the sign of the permutation.

10.38 Permuting the columns of a matrix

Suppose $A = \begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,n} \end{pmatrix}$ is an n-by-n matrix and $(m_1, \dots, m_n)$ is a permutation. Then

$$\det\begin{pmatrix} A_{\cdot,m_1} & \dots & A_{\cdot,m_n} \end{pmatrix} = \bigl(\operatorname{sign}(m_1, \dots, m_n)\bigr) \det A.$$

The next result about determinants will also be useful.

10.39 Determinant is a linear function of each column

Suppose k, n are positive integers with $1 \le k \le n$. Fix n-by-1 matrices $A_{\cdot,1}, \dots, A_{\cdot,n}$ except $A_{\cdot,k}$. Then the function that takes an n-by-1 column vector $A_{\cdot,k}$ to

$$\det\begin{pmatrix} A_{\cdot,1} & \dots & A_{\cdot,k} & \dots & A_{\cdot,n} \end{pmatrix}$$

is a linear map from the vector space of n-by-1 matrices with entries in $\mathbf{F}$ to $\mathbf{F}$.

Now we are ready to prove one of the key properties about determinants of square matrices. This property will enable us to connect the determinant of an operator with the determinant of its matrix. Note that this proof is considerably more complicated than the proof of the corresponding result about the trace (see 10.14).

10.40 Determinant is multiplicative

Suppose A and B are square matrices of the same size. Then

$$\det(AB) = \det(BA) = (\det A)(\det B).$$

Now we can prove that the determinant of the matrix of an operator is independent of the basis with respect to which the matrix is computed.

10.41 Determinant of matrix of operator does not depend on basis

Let $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then

$$\det \mathcal{M}\bigl(T, (u_1, \dots, u_n)\bigr) = \det \mathcal{M}\bigl(T, (v_1, \dots, v_n)\bigr).$$

The result below states that the determinant of an operator equals the determinant of the matrix of the operator. This theorem does not specify a basis because, by the result above, the determinant of the matrix of an operator is the same for every choice of basis.

10.42 Determinant of an operator equals determinant of its matrix

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

If we know the matrix of an operator on a complex vector space, the result above allows us to find the product of all the eigenvalues without finding any of the eigenvalues.

We can use 10.42 to give easy proofs of some useful properties about determinants of operators by shifting to the language of determinants of matrices, where certain properties have already been proved or are obvious. We carry out this procedure in the next result.

10.44 Determinant is multiplicative

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

$$\det(ST) = \det(TS) = (\det S)(\det T).$$

The Sign of the Determinant

We proved the basic results of linear algebra before introducing determinants in this final chapter. Although determinants have value as a research tool in more advanced subjects, they play little role in basic linear algebra (when the subject is done right).

Determinants do have one important application in undergraduate mathematics, namely, in computing certain volumes and integrals. In this subsection we interpret the meaning of the sign of the determinant on a real vector space. Then in the final subsection we will use the linear algebra we have learned to make clear the connection between determinants and these applications. Thus we will be dealing with a part of analysis that uses linear algebra.

We will begin with some purely linear algebra results that will also be useful when investigating volumes. Our setting will be inner product spaces. Recall that an isometry on an inner product space is an operator that preserves norms. The next result shows that every isometry has determinant with absolute value 1.

10.45 Isometries have determinant with absolute value 1

Suppose V is an inner product space and $S \in \mathcal{L}(V)$ is an isometry. Then $|\det S| = 1$.

The Real Spectral Theorem 7.29 states that a self-adjoint operator T on a real inner product space has an orthonormal basis consisting of eigenvectors. With respect to such a basis, the number of times each eigenvalue appears on the diagonal of $\mathcal{M}(T)$ is its multiplicity. Thus $\det T$ equals the product of its eigenvalues, counting multiplicity (of course, this holds for every operator, self-adjoint or not, on a complex vector space).

Recall that if V is an inner product space and $T \in \mathcal{L}(V)$, then T^*T is a positive operator and hence has a unique positive square root, denoted $\sqrt{T^*T}$ (see 7.35 and 7.36). Because $\sqrt{T^*T}$ is positive, all its eigenvalues are nonnegative (again, see 7.35), and hence $\det \sqrt{T^*T} \geq 0$.

Volume

The next result will be a key tool in our investigation of volume.

10.47 $|\det T| = \det \sqrt{T^*T}$

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

$$|\det T| = \det \sqrt{T^*T}.$$

Now we turn to the question of volume in $\mathbf{R}^n$. Fix a positive integer n for the rest of this subsection. We will consider only the real inner product space $\mathbf{R}^n$, with its standard inner product.

We would like to assign to each subset Ω of $\mathbf{R}^n$ its n-dimensional volume (when $n = 2$, this is usually called area instead of volume). We begin with boxes, where we have a good intuitive notion of volume.

10.48 Definition ***box***

A ***box*** in $\mathbf{R}^n$ is a set of the form

$$\{(y_1, \dots, y_n) \in \mathbf{R}^n : x_j < y_j < x_j + r_j \text{ for } j = 1, \dots, n\},$$

where $r_1, \dots, r_n$ are positive numbers and $(x_1, \dots, x_n) \in \mathbf{R}^n$. The numbers $r_1, \dots, r_n$ are called the ***side lengths*** of the box.

You should verify that when $n = 2$, a box is a rectangle with sides parallel to the coordinate axes, and that when $n = 3$, a box is a familiar 3-dimensional box with sides parallel to the coordinate axes.

The next definition fits with our intuitive notion of volume, because we define the volume of a box to be the product of the side lengths of the box.

10.49 Definition ***volume of a box***

The ***volume*** of a box B in $\mathbf{R}^n$ with side lengths $r_1, \dots, r_n$ is defined to be $r_1 \cdots r_n$ and is denoted by volume B.

To define the volume of an arbitrary set $\Omega \subset \mathbf{R}^n$, the idea is to write Ω as a subset of a union of many small boxes, then add up the volumes of these small boxes. As we approximate Ω more accurately by unions of small boxes, we get a better estimate of volume Ω.

10.50 Definition ***volume***

Suppose $\Omega \subset \mathbf{R}^n$. Then the ***volume*** of Ω, denoted volume Ω, is defined to be the infimum of

$$\text{volume } B_1 + \text{volume } B_2 + \cdots,$$

where the infimum is taken over all sequences $B_1, B_2, \dots$ of boxes in $\mathbf{R}^n$ whose union contains Ω.

We will work only with an intuitive notion of volume. Our purpose in this book is to understand linear algebra, whereas notions of volume belong to analysis (although volume is intimately connected with determinants, as we will soon see). Thus for the rest of this section we will rely on intuitive notions of volume rather than on a rigorous development, although we shall maintain our usual rigor in the linear algebra parts of what follows. Everything said here about volume will be correct if appropriately interpreted—the intuitive approach used here can be converted into appropriate correct definitions, correct statements, and correct proofs using the machinery of analysis.

10.51 Notation $T(\Omega)$

For T a function defined on a set Ω, define $T(\Omega)$ by

$$T(\Omega) = \{Tx : x \in \Omega\}.$$

For $T \in \mathcal{L}(\mathbf{R}^n)$ and $\Omega \subset \mathbf{R}^n$, we seek a formula for volume $T(\Omega)$ in terms of T and volume Ω. We begin by looking at positive operators.

10.52 Positive operators change volume by factor of determinant

Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ is a positive operator and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume } T(\Omega) = (\det T)(\text{volume } \Omega).$$

Our next tool is the following result, which states that isometries do not change volume.

10.53 An isometry does not change volume

Suppose $S \in \mathcal{L}(\mathbf{R}^n)$ is an isometry and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume}\, S(\Omega) = \text{volume}\, \Omega.$$

Now we can prove that an operator $T \in \mathcal{L}(\mathbf{R}^n)$ changes volume by a factor of $|\det T|$. Note the huge importance of the Polar Decomposition in the proof.

10.54 T changes volume by factor of $|\det T|$

Suppose $T \in \mathcal{L}(\mathbf{R}^n)$ and $\Omega \subset \mathbf{R}^n$. Then

$$\text{volume}\, T(\Omega) = |\det T|(\text{volume}\, \Omega).$$

The result that we just proved leads to the appearance of determinants in the formula for change of variables in multivariable integration. To describe this, we will again be vague and intuitive.

Throughout this book, almost all the functions we have encountered have been linear. Thus please be aware that the functions f and σ in the material below are not assumed to be linear.

The next definition aims at conveying the idea of the integral; it is not intended as a rigorous definition.

10.55 Definition ***integral, $\int_\Omega f$***

If $\Omega \subset \mathbf{R}^n$ and f is a real-valued function on Ω, then the ***integral*** of f over Ω, denoted $\int_\Omega f$ or $\int_\Omega f(x)\,dx$, is defined by breaking Ω into pieces small enough that f is almost constant on each piece. On each piece, multiply the (almost constant) value of f by the volume of the piece, then add up these numbers for all the pieces, getting an approximation to the integral that becomes more accurate as Ω is divided into finer pieces.

Actually, Ω in the definition above needs to be a reasonable set (for example, open or measurable) and f needs to be a reasonable function (for example, continuous or measurable), but we will not worry about those technicalities. Also, notice that the x in $\int_\Omega f(x)\,dx$ is a dummy variable and could be replaced with any other symbol.

Now we define the notions of differentiable and derivative. Notice that in this context, the derivative is an operator, not a number as in one-variable calculus.

10.56 Definition ***differentiable, derivative,*** $\sigma'(x)$

Suppose Ω is an open subset of $\mathbf{R}^n$ and σ is a function from Ω to $\mathbf{R}^n$. For $x \in \Omega$, the function σ is called ***differentiable*** at x if there exists an operator $T \in \mathcal{L}(\mathbf{R}^n)$ such that

$$\lim_{y \to 0} \frac{\|\sigma(x+y) - \sigma(x) - Ty\|}{\|y\|} = 0.$$

If σ is differentiable at x, then the unique operator $T \in \mathcal{L}(\mathbf{R}^n)$ satisfying the equation above is called the ***derivative*** of σ at x and is denoted by $\sigma'(x)$.

The idea of the derivative is that for x fixed and $\|y\|$ small,

$$\sigma(x+y) \approx \sigma(x) + \big(\sigma'(x)\big)(y);$$

because $\sigma'(x) \in \mathcal{L}(\mathbf{R}^n)$, this makes sense.

Suppose Ω is an open subset of $\mathbf{R}^n$ and σ is a function from Ω to $\mathbf{R}^n$. We can write

$$\sigma(x) = \big(\sigma_1(x), \dots, \sigma_n(x)\big),$$

where each σ_j is a function from Ω to $\mathbf{R}$. The partial derivative of σ_j with respect to the k^{th} coordinate is denoted $D_k\sigma_j$. Evaluating this partial derivative at a point $x \in \Omega$ gives $D_k\sigma_j(x)$. If σ is differentiable at x, then the matrix of $\sigma'(x)$ with respect to the standard basis of $\mathbf{R}^n$ contains $D_k\sigma_j(x)$ in row j, column k (this is left as an exercise). In other words,

10.57
$$\mathcal{M}\big(\sigma'(x)\big) = \begin{pmatrix} D_1\sigma_1(x) & \dots & D_n\sigma_1(x) \\ \vdots & & \vdots \\ D_1\sigma_n(x) & \dots & D_n\sigma_n(x) \end{pmatrix}.$$

Now we can state the change of variables integration formula. Some additional mild hypotheses are needed for f and σ' (such as continuity or measurability), but we will not worry about them because the proof below is really a pseudoproof that is intended to convey the reason the result is true.

The result below is called a change of variables formula because you can think of $y = \sigma(x)$ as a change of variables.

10.58 Change of variables in an integral

Suppose Ω is an open subset of $\mathbf{R}^n$ and $\sigma : \Omega \to \mathbf{R}^n$ is differentiable at every point of Ω. If f is a real-valued function defined on $\sigma(\Omega)$, then

$$\int_{\sigma(\Omega)} f(y)\,dy = \int_{\Omega} f\bigl(\sigma(x)\bigr)\lvert\det \sigma'(x)\rvert\,dx.$$

The key point when making a change of variables is that the factor of $\lvert\det \sigma'(x)\rvert$ must be included when making a substitution $y = f(x)$, as in the right side of 10.58.

Photo Credits

- page 1: Pierre Louis Dumesnil; 1884 copy by Nils Forsberg/Public domain image from *Wikimedia*.
- page 14: George M. Bergman/Archives of the Mathematisches Forschungsinstitut Oberwolfach.
- page 23: Gottlieb Biermann; photo by A. Wittmann/Public domain image from *Wikimedia*.
- page 49: Mostafa Azizi/Public domain image from *Wikimedia*.
- page 57: Hans-Peter Postel/Public domain image from *Wikimedia*.
- page 69: Public domain image from *Wikimedia*.
- page 84: Public domain image from *Wikimedia*. Original painting is in Tate Britain.
- page 97: Public domain image from *Wikimedia*.
- page 109: Public domain image from *Wikimedia*. Original fresco is in the Vatican.
- page 118: Public domain image from *Wikimedia*.

Index