1 Linear Algebra

Basic Object:	Vector Spaces
Basic Map:	Linear Transformations
Basic Goal:	Equivalences for the Invertibility of Matrices

1.1 Introduction

Though a bit of an exaggeration, it can be said that a mathematical problem can be solved only if it can be reduced to a calculation in linear algebra. And a calculation in linear algebra will reduce ultimately to the solving of a system of linear equations, which in turn comes down to the manipulation of matrices. Throughout this text and, more importantly, throughout mathematics, linear algebra is a key tool (or more accurately, a collection of intertwining tools) that is critical for doing calculations.

The power of linear algebra lies not only in our ability to manipulate matrices in order to solve systems of linear equations. The abstraction of these concrete objects to the ideas of vector spaces and linear transformations allows us to see the common conceptual links between many seemingly disparate subjects. (Of course, this is the advantage of any good abstraction.) For example, the study of solutions to linear differential equations has, in part, the same feel as trying to model the hood of a car with cubic polynomials, since both the space of solutions to a linear differential equation and the space of cubic polynomials that model a car hood form vector spaces.

The key theorem of linear algebra, discussed in Section 1.6, gives many equivalent ways of telling when a system of n linear equations in n unknowns has a solution. Each of the equivalent conditions is important. What is remarkable and what gives linear algebra its oomph is that they are all the same.

1.2 The Basic Vector Space $\mathbb{R}^n$

The quintessential vector space is $\mathbb{R}^n$, the set of all n-tuples of real numbers

$$\{(x_1, \ldots, x_n) : x_i \in \mathbb{R}^n\}.$$

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

As we will see in the next section, what makes this a vector space is that we can add together two n-tuples to get another n-tuple

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and that we can multiply each n-tuple by a real number λ

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

to get another n-tuple. Of course each n-tuple is usually called a vector and the real numbers λ are called scalars. When $n = 2$ and when $n = 3$ all of this reduces to the vectors in the plane and in space that most of us learned in high school.

The natural map from some $\mathbb{R}^n$ to an $\mathbb{R}^m$ is given by matrix multiplication. Write a vector $\mathbf{x} \in \mathbb{R}^n$ as a column vector:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Similarly, we can write a vector in $\mathbb{R}^m$ as a column vector with m entries. Let A be an $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{pmatrix}.$$

Then $A\mathbf{x}$ is the m-tuple:

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \ldots & \ldots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}.$$

For any two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and any two scalars λ and μ, we have

$$A(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda A\mathbf{x} + \mu A\mathbf{y}.$$

In the next section we will use the linearity of matrix multiplication to motivate the definition for a linear transformation between vector spaces.

Now to relate all of this to the solving of a system of linear equations. Suppose we are given numbers $b_1, \ldots, b_m$ and numbers $a_{11}, \ldots, a_{mn}$. Our goal is to find n numbers $x_1, \ldots, x_n$ that solve the following system of linear equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Calculations in linear algebra will frequently reduce to solving a system of linear equations. When there are only a few equations, we can find the solutions by hand, but as the number of equations increases, the calculations quickly turn from enjoyable algebraic manipulations into nightmares of notation. These nightmarish complications arise not from any single theoretical difficulty but instead stem solely from trying to keep track of the many individual minor details. In other words, it is a problem in bookkeeping.

Write

$$\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix}$$

and our unknowns as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Then we can rewrite our system of linear equations in the more visually appealing form of

$$A\mathbf{x} = \mathbf{b}.$$

When $m > n$ (when there are more equations than unknowns), we expect there to be, in general, no solutions. For example, when $m = 3$ and $n = 2$, this corresponds geometrically to the fact that three lines in a plane will usually have no common point of intersection. When $m < n$ (when there are more unknowns than equations), we expect there to be, in general, many solutions. In the case when $m = 2$ and $n = 3$, this corresponds geometrically to the fact that two planes in space will usually intersect in an entire line. Much of the machinery of linear algebra deals with the remaining case when $m = n$.

Thus we want to find the $n \times 1$ column vector $\mathbf{x}$ that solves $A\mathbf{x} = \mathbf{b}$, where A is a given $n \times n$ matrix and $\mathbf{b}$ is a given $n \times 1$ column vector. Suppose that the square matrix A has an inverse matrix A^{-1} (which means that A^{-1} is also $n \times n$ and more importantly that $A^{-1}A = I$, with I the identity matrix). Then our solution will be

$$\mathbf{x} = A^{-1}\mathbf{b}$$

since

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = I\mathbf{b} = \mathbf{b}.$$

Thus solving our system of linear equations comes down to understanding when the $n \times n$ matrix A has an inverse. (If an inverse matrix exists, then there are algorithms for its calculation.)

The key theorem of linear algebra, stated in Section 1.6, is in essence a list of many equivalences for when an $n \times n$ matrix has an inverse. It is thus essential to understanding when a system of linear equations can be solved.

1.3 Vector Spaces and Linear Transformations

The abstract approach to studying systems of linear equations starts with the notion of a vector space.

Definition 1.3.1 A set V is a *vector space* over the real numbers[1] $\mathbb{R}$ if there are maps:

1. $\mathbb{R} \times V \to V$, denoted by $a \cdot v$ or av for all real numbers a and elements v in V,
2. $V \times V \to V$, denoted by $v + w$ for all elements v and w in the vector space V,

with the following properties.

(a) There is an element 0, in V such that $0 + v = v$ for all $v \in V$.
(b) For each $v \in V$, there is an element $(-v) \in V$ with $v + (-v) = 0$.
(c) For all $v, w \in V$, $v + w = w + v$.
(d) For all $a \in \mathbb{R}$ and for all $v, w \in V$, we have that $a(v + w) = av + aw$.
(e) For all $a, b \in \mathbb{R}$ and all $v \in V$, $a(bv) = (a \cdot b)v$.
(f) For all $a, b \in \mathbb{R}$ and all $v \in V$, $(a + b)v = av + bv$.
(g) For all $v \in V$, $1 \cdot v = v$.

As a matter of notation, and to agree with common usage, the elements of a vector space are called *vectors* and the elements of $\mathbb{R}$ (or whatever field is being used) are called *scalars*. Note that the space $\mathbb{R}^n$ given in the last section certainly satisfies these conditions.

The natural map between vector spaces is that of a linear transformation.

Definition 1.3.2 A *linear transformation* $T : V \to W$ is a function from a vector space V to a vector space W such that for any real numbers a_1 and a_2 and any vectors v_1 and v_2 in V, we have

$$T(a_1 v_1 + a_2 v_2) = a_1 T(v_1) + a_2 T(v_2).$$

Matrix multiplication from an $\mathbb{R}^n$ to an $\mathbb{R}^m$ gives an example of a linear transformation.

Definition 1.3.3 A subset U of a vector space V is a *subspace* of V if U is itself a vector space.

[1] The real numbers can be replaced by the complex numbers and in fact by any field (which will be defined in Chapter 11 on algebra).

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

In practice, it is usually easy to see if a subset of a vector space is in fact a subspace, by the following proposition, whose proof is left to the reader.

Proposition 1.3.4 *A subset U of a vector space V is a subspace of V if U is closed under addition and scalar multiplication.*

Given a linear transformation $T: V \to W$, there are naturally occurring subspaces of both V and W.

Definition 1.3.5 If $T: V \to W$ is a linear transformation, then the *kernel* of T is:

$$\ker(T) = \{v \in V : T(v) = 0\}$$

and the *image* of T is

$$\mathrm{Im}(T) = \{w \in W : \text{there exists a } v \in V \text{ with } T(v) = w\}.$$

The kernel is a subspace of V, since if v_1 and v_2 are two vectors in the kernel and if a and b are any two real numbers, then

$$\begin{aligned} T(av_1 + bv_2) &= aT(v_1) + bT(v_2) \\ &= a \cdot 0 + b \cdot 0 \\ &= 0. \end{aligned}$$

In a similar way we can show that the image of T is a subspace of W, which we leave for one of the exercises.

If the only vector spaces that ever occurred were column vectors in $\mathbb{R}^n$, then even this mild level of abstraction would be silly. This is not the case. Here we look at only one example. Let $C^k[0,1]$ be the set of all real-valued functions with domain the unit interval $[0,1]$:

$$f: [0,1] \to \mathbb{R}$$

such that the kth derivative of f exists and is continuous. Since the sum of any two such functions and a multiple of any such function by a scalar will still be in $C^k[0,1]$, we have a vector space. Though we will officially define dimension in the next section, $C^k[0,1]$ will be infinite dimensional (and thus definitely not some $\mathbb{R}^n$). We can view the derivative as a linear transformation from $C^k[0,1]$ to those functions with one less derivative, $C^{k-1}[0,1]$:

$$\frac{\mathrm{d}}{\mathrm{d}x}: C^k[0,1] \to C^{k-1}[0,1].$$

The kernel of $\frac{\mathrm{d}}{\mathrm{d}x}$ consists of those functions with $\frac{\mathrm{d}f}{\mathrm{d}x} = 0$, namely constant functions.

Now consider the differential equation

$$\frac{\mathrm{d}^2 f}{\mathrm{d}x^2} + 3\frac{\mathrm{d}f}{\mathrm{d}x} + 2f = 0.$$

Let T be the linear transformation:

$$T = \frac{\mathrm{d}^2}{\mathrm{d}x^2} + 3\frac{\mathrm{d}}{\mathrm{d}x} + 2I : C^2[0,1] \to C^0[0,1].$$

The problem of finding a solution $f(x)$ to the original differential equation can now be translated to finding an element of the kernel of T. This suggests the possibility (which indeed is true) that the language of linear algebra can be used to understand solutions to (linear) differential equations.

1.4 Bases, Dimension, and Linear Transformations as Matrices

Our next goal is to define the dimension of a vector space.

Definition 1.4.1 A set of vectors $(v_1, \ldots, v_n)$ form a *basis* for the vector space V if given any vector v in V, there are unique scalars $a_1, \ldots, a_n \in \mathbb{R}$ with $v = a_1 v_1 + \cdots + a_n v_n$.

Definition 1.4.2 The *dimension* of a vector space V, denoted by $\dim(V)$, is the number of elements in a basis.

As it is far from obvious that the number of elements in a basis will always be the same, no matter which basis is chosen, in order to make the definition of the dimension of a vector space well defined we need the following theorem (which we will not prove).

Theorem 1.4.3 *All bases of a vector space V have the same number of elements.*

For $\mathbb{R}^n$, the usual basis is

$$\{(1,0,\ldots,0),(0,1,0,\ldots,0),\ldots,(0,\ldots,0,1)\}.$$

Thus $\mathbb{R}^n$ is n-dimensional. Of course if this were not true, the above definition of dimension would be wrong and we would need another. This is an example of the principle mentioned in the introduction. We have a good intuitive understanding of what dimension should mean for certain specific examples: a line needs to be one dimensional, a plane two dimensional and space three dimensional. We then come up with a sharp definition. If this definition gives the “correct” answer for our three already understood examples, we are somewhat confident that the definition has

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

indeed captured what is meant by, in this case, dimension. Then we can apply the definition to examples where our intuitions fail.

Linked to the idea of a basis is the idea of linear independence.

Definition 1.4.4 Vectors $(v_1, \ldots, v_n)$ in a vector space V are *linearly independent* if whenever

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

it must be the case that the scalars $a_1, \ldots, a_n$ must all be zero.

Intuitively, a collection of vectors are linearly independent if they all point in different directions. A basis consists then in a collection of linearly independent vectors that span the vector space, where by span we mean the following.

Definition 1.4.5 A set of vectors $(v_1, \ldots, v_n)$ *span* the vector space V if given any vector v in V, there are scalars $a_1, \ldots, a_n \in \mathbb{R}$ with $v = a_1 v_1 + \cdots + a_n v_n$.

Our goal now is to show how all linear transformations $T : V \to W$ between finite-dimensional spaces can be represented as matrix multiplication, provided we fix bases for the vector spaces V and W.

First fix a basis $\{v_1, \ldots, v_n\}$ for V and a basis $\{w_1, \ldots, w_m\}$ for W. Before looking at the linear transformation T, we need to show how each element of the n-dimensional space V can be represented as a column vector in $\mathbb{R}^n$ and how each element of the m-dimensional space W can be represented as a column vector of $\mathbb{R}^n$. Given any vector v in V, by the definition of basis, there are unique real numbers $a_1, \ldots, a_n$ with

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

We thus represent the vector v with the column vector:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, for any vector w in W, there are unique real numbers $b_1, \ldots, b_m$ with

$$w = b_1 w_1 + \cdots + b_m w_m.$$

Here we represent w as the column vector

$$\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Note that we have established a correspondence between vectors in V and W and column vectors $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. More technically, we can show that V is isomorphic to $\mathbb{R}^n$ (meaning that there is a one-to-one, onto linear transformation from V to $\mathbb{R}^n$ and that W is isomorphic to $\mathbb{R}^m$, though it must be emphasized that the actual correspondence only exists after a basis has been chosen (which means that while the isomorphism exists, it is not canonical; this is actually a big deal, as in practice it is unfortunately often the case that no basis is given to us).

We now want to represent a linear transformation $T: V \to W$ as an $m \times n$ matrix A. For each basis vector v_i in the vector space V, $T(v_i)$ will be a vector in W. Thus there will exist real numbers $a_{1i}, \dots, a_{mi}$ such that

$$T(v_i) = a_{1i}w_1 + \cdots + a_{mi}w_m.$$

We want to see that the linear transformation T will correspond to the $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix}.$$

Given any vector v in V, with $v = a_1v_1 + \cdots + a_nv_n$, we have

$$\begin{aligned} T(v) &= T(a_1v_1 + \cdots + a_nv_n) \\ &= a_1T(v_1) + \cdots + a_nT(v_n) \\ &= a_1(a_{11}w_1 + \cdots + a_{m1}w_m) + \cdots \\ &\quad + a_n(a_{1n}w_1 + \cdots + a_{mn}w_m). \end{aligned}$$

But under the correspondences of the vector spaces with the various column spaces, this can be seen to correspond to the matrix multiplication of A times the column vector corresponding to the vector v:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \dots & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Note that if $T: V \to V$ is a linear transformation from a vector space to itself, then the corresponding matrix will be $n \times n$, a square matrix.

Given different bases for the vector spaces V and W, the matrix associated to the linear transformation T will change. A natural problem is to determine when two matrices actually represent the same linear transformation, but under different bases. This will be the goal of Section 1.7.

1.5 The Determinant

Our next task is to give a definition for the determinant of a matrix. In fact, we will give three alternative descriptions of the determinant. All three are equivalent; each has its own advantages.

Our first method is to define the determinant of a 1×1 matrix and then to define recursively the determinant of an $n \times n$ matrix.

Since 1×1 matrices are just numbers, the following should not at all be surprising.

Definition 1.5.1 The *determinant* of a 1×1 matrix (a) is the real-valued function

$$\det(a) = a.$$

This should not yet seem significant.

Before giving the definition of the determinant for a general $n \times n$ matrix, we need a little notation. For an $n \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{pmatrix},$$

denote by A_{ij} the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and the jth column. For example, if $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, then $A_{12} = (a_{21})$. Similarly if $A = \begin{pmatrix} 2 & 3 & 5 \\ 6 & 4 & 9 \\ 7 & 1 & 8 \end{pmatrix}$, then $A_{12} = \begin{pmatrix} 6 & 9 \\ 7 & 8 \end{pmatrix}$.

Since we have a definition for the determinant for 1×1 matrices, we will now assume by induction that we know the determinant of any $(n-1) \times (n-1)$ matrix and use this to find the determinant of an $n \times n$ matrix.

Definition 1.5.2 Let A be an $n \times n$ matrix. Then the *determinant* of A is

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+1} a_{1k} \det(A_{1k}).$$

Thus for $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, we have

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) = a_{11}a_{22} - a_{12}a_{21},$$

which is what most of us think of as the determinant. The determinant of our above 3×3 matrix is:

$$\det \begin{pmatrix} 2 & 3 & 5 \\ 6 & 4 & 9 \\ 7 & 1 & 8 \end{pmatrix} = 2 \det \begin{pmatrix} 4 & 9 \\ 1 & 8 \end{pmatrix} - 3 \det \begin{pmatrix} 6 & 9 \\ 7 & 8 \end{pmatrix} + 5 \det \begin{pmatrix} 6 & 4 \\ 7 & 1 \end{pmatrix}.$$

While this definition is indeed an efficient means to describe the determinant, it obscures most of the determinant's uses and intuitions.

The second way we can describe the determinant has built into it the key algebraic properties of the determinant. It highlights function-theoretic properties of the determinant.

Denote the $n \times n$ matrix A as $A = (A_1, \ldots, A_n)$, where A_i denotes the ith column:

$$A_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}.$$

Definition 1.5.3 The *determinant* of A is defined as the unique real-valued function

$$\det\colon \textit{Matrices} \to \mathbb{R}$$

satisfying:

(a) $\det(A_1, \ldots, \lambda A_k, \ldots, A_n) = \lambda \det(A_1, \ldots, A_k)$.
(b) $\det(A_1, \ldots, A_k + \lambda A_i, \ldots, A_n) = \det(A_1, \ldots, A_n)$ for $k \neq i$.
(c) det(Identity matrix) = 1.

Thus, treating each column vector of a matrix as a vector in $\mathbb{R}^n$, the determinant can be viewed as a special type of function from $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ to the real numbers.

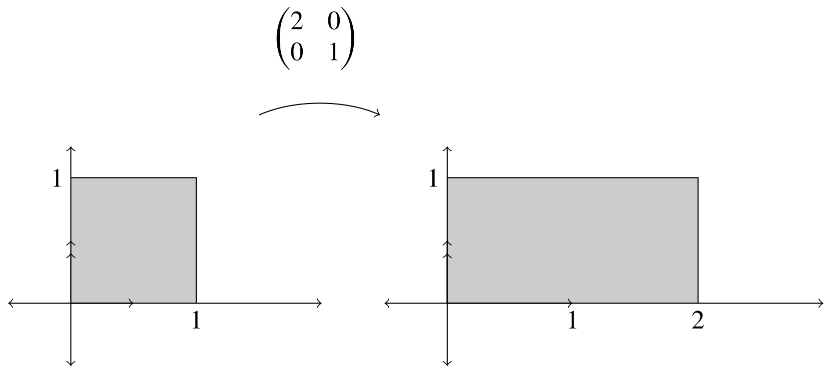
In order to be able to use this definition, we would have to prove that such a function on the space of matrices, satisfying conditions (a) through (c), even exists and then that it is unique. Existence can be shown by checking that our first (inductive) definition for the determinant satisfies these conditions, though it is a painful calculation. The proof of uniqueness can be found in almost any linear algebra text, and comes down to using either elementary column operations or elementary row operations.

The third definition for the determinant is the most geometric but is also the most vague. We must think of an $n \times n$ matrix A as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$. Then A will map the unit cube in $\mathbb{R}^n$ to some different object (a parallelepiped). The unit cube has, by definition, a volume of one.

Definition 1.5.4 The *determinant* of the matrix A is the signed volume of the image of the unit cube.

This is not well defined, as the very method of defining the volume of the image has not been described. In fact, most would define the signed volume of the image to be the number given by the determinant using one of the two earlier definitions. But this can all be made rigorous, though at the price of losing much of the geometric insight.

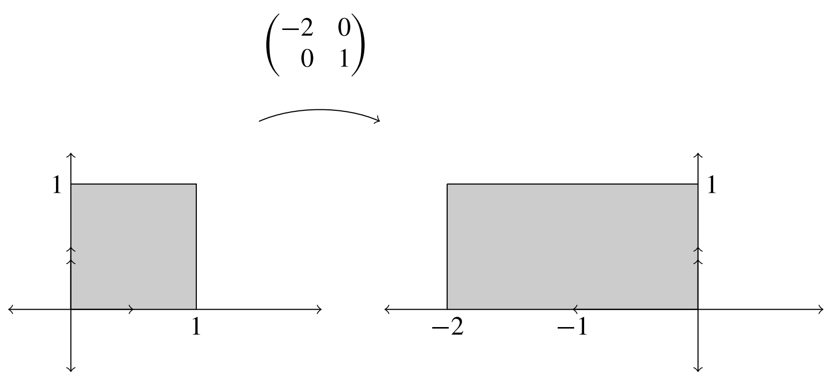
Let us look at some examples: the matrix $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ takes the unit square to



Since the area is doubled, we must have

$$\det(A) = 2.$$

Signed volume means that if the orientations of the edges of the unit cube are changed, then we must have a negative sign in front of the volume. For example, consider the matrix $A = \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}$. Here the image is



https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Note that the orientations of the sides are flipped. Since the area is still doubled, the definition will force

$$\det(A) = -2.$$

To rigorously define orientation is somewhat tricky (we do it in Chapter 6), but its meaning is straightforward.

The determinant has many algebraic properties.

Lemma 1.5.5 *If A and B are $n \times n$ matrices, then*

$$\det(AB) = \det(A)\det(B).$$

This can be proven either by a long calculation or by concentrating on the definition of the determinant as the change of volume of a unit cube.

1.6 The Key Theorem of Linear Algebra

Here is the key theorem of linear algebra. (Note: we have yet to define eigenvalues and eigenvectors, but we will in Section 1.8.)

Theorem 1.6.1 (Key Theorem) *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. *A is invertible.*
2. $\det(A) \neq 0$.
3. $\ker(A) = 0$.
4. *If $\mathbf{b}$ is a column vector in $\mathbb{R}^n$, there is a unique column vector $\mathbf{x}$ in $\mathbb{R}^n$ satisfying $A\mathbf{x} = \mathbf{b}$.*
5. *The columns of A are linearly independent $n \times 1$ column vectors.*
6. *The rows of A are linearly independent $1 \times n$ row vectors.*
7. *The transpose A^t of A is invertible. (Here, if $A = (a_{ij})$, then $A^t = (a_{ji})$).*
8. *All of the eigenvalues of A are non-zero.*

We can restate this theorem in terms of linear transformations.

Theorem 1.6.2 (Key Theorem) *Let $T : V \to V$ be a linear transformation. Then the following are equivalent.*

1. *T is invertible.*
2. $\det(T) \neq 0$, *where the determinant is defined by a choice of basis on V.*
3. $\ker(T) = 0$.
4. *If b is a vector in V, there is a unique vector v in V satisfying $T(v) = b$.*

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

5. *For any basis $v_1, \ldots, v_n$ of V, the image vectors $T(v_1), \ldots, T(v_n)$ are linearly independent.*
6. *For any basis $v_1, \ldots, v_n$ of V, if S denotes the transpose linear transformation of T, then the image vectors $S(v_1), \ldots, S(v_n)$ are linearly independent.*
7. *The transpose of T is invertible. (Here the transpose is defined by a choice of basis on V.)*
8. *All of the eigenvalues of T are non-zero.*

In order to make the correspondence between the two theorems clear, we must worry about the fact that we only have definitions of the determinant and the transpose for matrices, not for linear transformations. While we do not show it, both notions can be extended to linear transformations, provided a basis is chosen. But note that while the actual value $\det(T)$ will depend on a fixed basis, the condition that $\det(T) \neq 0$ does not. Similar statements hold for conditions (6) and (7). A proof is the goal of exercise 8, where you are asked to find any linear algebra book and then fill in the proof. It is unlikely that the linear algebra book will have this result as it is stated here. The act of translating is in fact part of the purpose of making this an exercise.

Each of the equivalences is important. Each can be studied on its own merits. It is remarkable that they are the same.

1.7 Similar Matrices

Recall that given a basis for an n-dimensional vector space V, we can represent a linear transformation

$$T : V \to V$$

as an $n \times n$ matrix A. Unfortunately, if you choose a different basis for V, the matrix representing the linear transformation T will be quite different from the original matrix A. The goal of this section is to find a clean criterion for when two matrices actually represent the same linear transformation but under a different choice of bases.

Definition 1.7.1 Two $n \times n$ matrices A and B are *similar* if there is an invertible matrix C such that

$$A = C^{-1}BC.$$

We want to see that two matrices are similar precisely when they represent the same linear transformation. Choose two bases for the vector space V, say $\{v_1, \ldots, v_n\}$ (the v basis) and $\{w_1, \ldots, w_n\}$ (the w basis). Let A be the matrix representing the linear transformation T for the v basis and let B be the matrix

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

representing the linear transformation for the w basis. We want to construct the matrix C so that $A = C^{-1}BC$.

Recall that given the v basis, we can write each vector $z \in V$ as an $n \times 1$ column vector as follows: we know that there are unique scalars $a_1, \ldots, a_n$ with

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

We then write z, with respect to the v basis, as the column vector:

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Similarly, there are unique scalars $b_1, \ldots, b_n$ so that

$$z = b_1 w_1 + \cdots + b_n w_n,$$

meaning that with respect to the w basis, the vector z is the column vector:

$$\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

The desired matrix C will be the matrix such that

$$C \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

If $C = (c_{ij})$, then the entries c_{ij} are precisely the numbers which yield:

$$w_i = c_{i1} v_1 + \cdots + c_{in} v_n.$$

Then, for A and B to represent the same linear transformation, we need the diagram:

$$\begin{array}{ccccc} & \mathbb{R}^n & \overset{A}{\rightarrow} & \mathbb{R}^n & \\ C & \downarrow & & \downarrow & C \\ & \mathbb{R}^n & \underset{B}{\rightarrow} & \mathbb{R}^n & \end{array}$$

to commute, meaning that $CA = BC$ or

$$A = C^{-1}BC,$$

as desired.

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Determining when two matrices are similar is a type of result that shows up throughout math and physics. Regularly you must choose some coordinate system (some basis) in order to write down anything at all, but the underlying math or physics that you are interested in is independent of the initial choice. The key question becomes: what is preserved when the coordinate system is changed? Similar matrices allow us to start to understand these questions.

1.8 Eigenvalues and Eigenvectors

In the last section we saw that two matrices represent the same linear transformation, under different choices of bases, precisely when they are similar. This does not tell us, though, how to choose a basis for a vector space so that a linear transformation has a particularly decent matrix representation. For example, the diagonal matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

is similar to the matrix

$$B = \begin{pmatrix} -1 & -2 & -2 \\ 12 & 7 & 4 \\ -9 & -3 & 0 \end{pmatrix},$$

but all recognize the simplicity of A as compared to B. (By the way, it is not obvious that A and B are similar; I started with A, chose a non-singular matrix C and then computed $C^{-1}AC$ to get B. I did not just suddenly "see" that A and B are similar. No, I rigged it to be so.)

One of the purposes behind the following definitions for eigenvalues and eigenvectors is to give us tools for picking out good bases. There are, though, many other reasons to understand eigenvalues and eigenvectors.

Definition 1.8.1 Let $T : V \to V$ be a linear transformation. Then a non-zero vector $v \in V$ will be an *eigenvector* of T with *eigenvalue* λ, a scalar, if

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

For an $n \times n$ matrix A, a non-zero column vector $\mathbf{x} \in \mathbb{R}^n$ will be an *eigenvector* with *eigenvalue* λ, a scalar, if

$$A\mathbf{x} = \lambda \mathbf{x}.$$

Geometrically, a vector v is an eigenvector of the linear transformation T with eigenvalue λ if T stretches v by a factor of λ.

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

For example,

$$\begin{pmatrix} -2 & -2 \\ 6 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -2 \end{pmatrix},$$

and thus 2 is an eigenvalue and $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ an eigenvector for the linear transformation represented by the 2×2 matrix $\begin{pmatrix} -2 & -2 \\ 6 & 5 \end{pmatrix}$.

Luckily there is an easy way to describe the eigenvalues of a square matrix, which will allow us to see that the eigenvalues of a matrix are preserved under a similarity transformation.

Proposition 1.8.2 *A number λ will be an eigenvalue of a square matrix A if and only if λ is a root of the polynomial*

$$P(t) = \det(tI - A).$$

The polynomial $P(t) = \det(tI - A)$ is called the *characteristic polynomial* of the matrix A.

Proof: Suppose that λ is an eigenvalue of A, with eigenvector v. Then $Av = \lambda v$, or

$$\lambda v - Av = 0,$$

where the zero on the right-hand side is the zero column vector. Then, putting in the identity matrix I, we have

$$0 = \lambda v - Av = (\lambda I - A)v.$$

Thus the matrix $\lambda I - A$ has a non-trivial kernel, v. By the key theorem of linear algebra, this happens precisely when

$$\det(\lambda I - A) = 0,$$

which means that λ is a root of the characteristic polynomial $P(t) = \det(tI - A)$. Since all of these directions can be reversed, we have our theorem. □

Theorem 1.8.3 *Let A and B be similar matrices. Then the characteristic polynomial of A is equal to the characteristic polynomial of B.*

---

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Proof: For A and B to be similar, there must be an invertible matrix C with $A = C^{-1}BC$. Then

$$\begin{aligned}\det(tI - A) &= \det(tI - C^{-1}BC)\\ &= \det(tC^{-1}C - C^{-1}BC)\\ &= \det(C^{-1})\det(tI - B)\det(C)\\ &= \det(tI - B)\end{aligned}$$

using that $1 = \det(C^{-1}C) = \det(C^{-1})\det(C)$. □

Since the characteristic polynomials for similar matrices are the same, this means that the eigenvalues must be the same.

Corollary 1.8.4 *The eigenvalues for similar matrices are equal.*

Thus to see if two matrices are similar, one can compute to see if the eigenvalues are equal. If they are not, the matrices are not similar. Unfortunately, in general, having equal eigenvalues does not force matrices to be similar. For example, the matrices

$$A = \begin{pmatrix} 2 & -7 \\ 0 & 2 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

both have eigenvalue 2 with multiplicity two, but they are not similar. (This can be shown by assuming that there is an invertible 2×2 matrix C with $C^{-1}AC = B$ and then showing that $\det(C) = 0$, contradicting the invertibility of C.)

Since the characteristic polynomial $P(t)$ does not change under a similarity transformation, the coefficients of $P(t)$ will also not change under a similarity transformation. But since the coefficients of $P(t)$ will themselves be (complicated) polynomials of the entries of the matrix A, we now have certain special polynomials of the entries of A that are invariant under a similarity transformation. One of these coefficients we have already seen in another guise, namely the determinant of A, as the following theorem shows. This theorem will more importantly link the eigenvalues of A to the determinant of A.

Theorem 1.8.5 *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues, counted with multiplicity, of a matrix A. Then*

$$\det(A) = \lambda_1 \cdots \lambda_n.$$

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Before proving this theorem, we need to discuss the idea of counting eigenvalues "with multiplicity." The difficulty is that a polynomial can have a root that must be counted more than once (e.g., the polynomial $(x-2)^2$ has the single root 2 which we want to count twice). This can happen in particular to the characteristic polynomial. For example, consider the matrix

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

which has as its characteristic polynomial the cubic

$$(t-5)(t-5)(t-4).$$

For the above theorem, we would list the eigenvalues as 4, 5, and 5, hence counting the eigenvalue 5 twice.

Proof: Since the eigenvalues $\lambda_1, \dots, \lambda_n$ are the (complex) roots of the characteristic polynomial $\det(tI - A)$, we have

$$(t-\lambda_1)\cdots(t-\lambda_n) = \det(tI - A).$$

Setting $t = 0$, we have

$$(-1)^n \lambda_1 \cdots \lambda_n = \det(-A).$$

In the matrix $(-A)$, each column of A is multiplied by (-1). Using the second definition of a determinant, we can factor out each of these (-1), to get

$$(-1)^n \lambda_1 \cdots \lambda_n = (-1)^n \det(A)$$

and our result. □

Now finally to turn back to determining a "good" basis for representing a linear transformation. The measure of "goodness" is how close the matrix is to being a diagonal matrix. We will restrict ourselves to a special, but quite prevalent, class: symmetric matrices. By symmetric, we mean that if $A = (a_{ij})$, then we require that the entry at the ith row and jth column (a_{ij}) must equal the entry at the jth row and the ith column (a_{ji}). Thus

$$\begin{pmatrix} 5 & 3 & 4 \\ 3 & 5 & 2 \\ 4 & 2 & 4 \end{pmatrix}$$

is symmetric but

$$\begin{pmatrix} 5 & 2 & 3 \\ 6 & 5 & 3 \\ 2 & 18 & 4 \end{pmatrix}$$

is not.

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

Theorem 1.8.6 *If A is a symmetric matrix, then there is a matrix B similar to A which is not only diagonal but has the entries along the diagonal being precisely the eigenvalues of A.*

Proof: The proof basically rests on showing that the eigenvectors for A form a basis in which A becomes our desired diagonal matrix. We will assume that the eigenvalues for A are distinct, as technical difficulties occur when there are eigenvalues with multiplicity.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be the eigenvectors for the matrix A, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Form the matrix

$$C = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n),$$

where the ith column of C is the column vector $\mathbf{v}_i$. We will show that the matrix $C^{-1}AC$ will satisfy our theorem. Thus we want to show that $C^{-1}AC$ equals the diagonal matrix

$$B = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Denote

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Then the above diagonal matrix B is the unique matrix with $B\mathbf{e}_i = \lambda_i \mathbf{e}_i$, for all i. Our choice for the matrix C now becomes clear as we observe that for all i, $C\mathbf{e}_i = \mathbf{v}_i$. Then we have

$$C^{-1}AC\mathbf{e}_i = C^{-1}A\mathbf{v}_i = C^{-1}(\lambda_i \mathbf{v}_i) = \lambda_i C^{-1}\mathbf{v}_i = \lambda_i \mathbf{e}_i,$$

giving us the theorem. □

This is of course not the end of the story. For non-symmetric matrices, there are other canonical ways of finding “good” similar matrices, such as the Jordan canonical form, the upper triangular form and rational canonical form.

1.9 Dual Vector Spaces

It pays to study functions. In fact, functions appear at times to be more basic than their domains. In the context of linear algebra, the natural class of functions is linear transformations, or linear maps from one vector space to another. Among all real vector spaces, there is one that seems simplest, namely the one-dimensional vector space of the real numbers $\mathbb{R}$. This leads us to examine a special type of linear transformation on a vector space, those that map the vector space to the real numbers, the set of which we will call the *dual space*. Dual spaces regularly show up in mathematics.

Let V be a vector space. The *dual vector space*, or *dual space*, is:

$$\begin{aligned} V^* &= \{\text{linear maps from } V \text{ to the real numbers } \mathbb{R}\} \\ &= \{v^* \colon V \to \mathbb{R} \mid v^* \text{ is linear}\}. \end{aligned}$$

One of the exercises is to show that the dual space V^* is itself a vector space.

Let $T\colon V \to W$ be a linear transformation. Then we can define a natural linear transformation

$$T^* \colon W^* \to V^*$$

from the dual of W to the dual of V as follows. Let $w^* \in W^*$. Then given any vector w in the vector space W, we know that $w^*(w)$ will be a real number. We need to define T^* so that $T^*(w^*) \in V^*$. Thus given any vector $v \in V$, we need $T^*(w^*)(v)$ to be a real number. Simply define

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

By the way, note that the direction of the linear transformation $T\colon V \to W$ is indeed reversed to $T^*\colon W^* \to V^*$. Also by "natural" we do not mean that the map T^* is "obvious" but instead that it can be uniquely associated to the original linear transformation T.

Such a dual map shows up in many different contexts. For example, if X and Y are topological spaces with a continuous map $F\colon X \to Y$ and if $C(X)$ and $C(Y)$ denote the sets of continuous real-valued functions on X and Y, then here the dual map

$$F^* \colon C(Y) \to C(X)$$

is defined by $F^*(g)(x) = g(F(x))$, where g is a continuous map on Y.

Attempts to abstractly characterize all such dual maps were a major theme of mid-twentieth-century mathematics and can be viewed as one of the beginnings of category theory.

1.10 Books

Mathematicians have been using linear algebra since they have been doing mathematics, but the styles, methods and terminologies have shifted. For example, if you look in a college course catalog in 1900, or probably even 1950, there will

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

be no undergraduate course called linear algebra. Instead there were courses such as "Theory of Equations" or simply "Algebra." As seen in one of the more popular textbooks in the first part of the twentieth century, Maxime Bocher's *Introduction to Higher Algebra* [18], the concern was on concretely solving systems of linear equations. The results were written in an algorithmic style. Modern-day computer programmers usually find this style of text far easier to understand than current math books. In the 1930s, a fundamental change in the way algebraic topics were taught occurred with the publication of Van der Waerden's *Modern Algebra* [192, 193], which was based on lectures of Emmy Noether and Emil Artin. Here a more abstract approach was taken. The first true modern-day linear algebra text, at least in English, was Halmos' *Finite-Dimensional Vector Spaces* [81]. Here the emphasis is on the idea of a vector space from the very beginning. Today there are many beginning texts. Some start with systems of linear equations and then deal with vector spaces, others reverse the process. A long-time favorite of many is Strang's *Linear Algebra and Its Applications* [185]. As a graduate student, you should volunteer to teach or assist teaching linear algebra as soon as possible.

Exercises

(**1**) Let $L\colon V \to W$ be a linear transformation between two vector spaces. Show that

$$\dim(\ker(L)) + \dim(\mathrm{Im}(L)) = \dim(V).$$

(**2**) Consider the set of all polynomials in one variable with real coefficients of degree less than or equal to three.
a. Show that this set forms a vector space of dimension four.
b. Find a basis for this vector space.
c. Show that differentiating a polynomial is a linear transformation.
d. Given the basis chosen in part (b), write down the matrix representative of the derivative.

(**3**) Let $T\colon V \to W$ be a linear transformation from a vector space V to a vector space W. Show that the image of T

$$\mathrm{Im}(T) = \{w \in W : \text{there exists a } v \in V \text{with } T(v) = w\}$$

is a subspace of W.

(**4**) Let A and B be two $n \times n$ invertible matrices. Prove that

$$(AB)^{-1} = B^{-1}A^{-1}.$$

(**5**) Let

$$A = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}.$$

Find a matrix C so that $C^{-1}AC$ is a diagonal matrix.

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press

(**6**) Denote the vector space of all functions

$$f : \mathbb{R} \to \mathbb{R}$$

which are infinitely differentiable by $C^\infty(\mathbb{R})$. This space is called the space of smooth functions.

a. Show that $C^\infty(\mathbb{R})$ is infinite dimensional.

b. Show that differentiation is a linear transformation:

$$\frac{\mathrm{d}}{\mathrm{d}x} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}).$$

c. For a real number λ, find an eigenvector for $\frac{\mathrm{d}}{\mathrm{d}x}$ with eigenvalue λ.

(**7**) Let V be a finite-dimensional vector space. Show that the dual vector space V^* has the same dimension as V.

(**8**) Find a linear algebra text. Use it to prove the key theorem of linear algebra. Note that this is a long exercise but is to be taken seriously.

(**9**) For a vector space V, show that the dual space V^* is also a vector space.

https://doi.org/10.1017/9781108992879.002 Published online by Cambridge University Press